Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer

Petralia, Francesca ; Tignor, Nicole ; Reva, Boris ; et al ; Nazarian, Javad

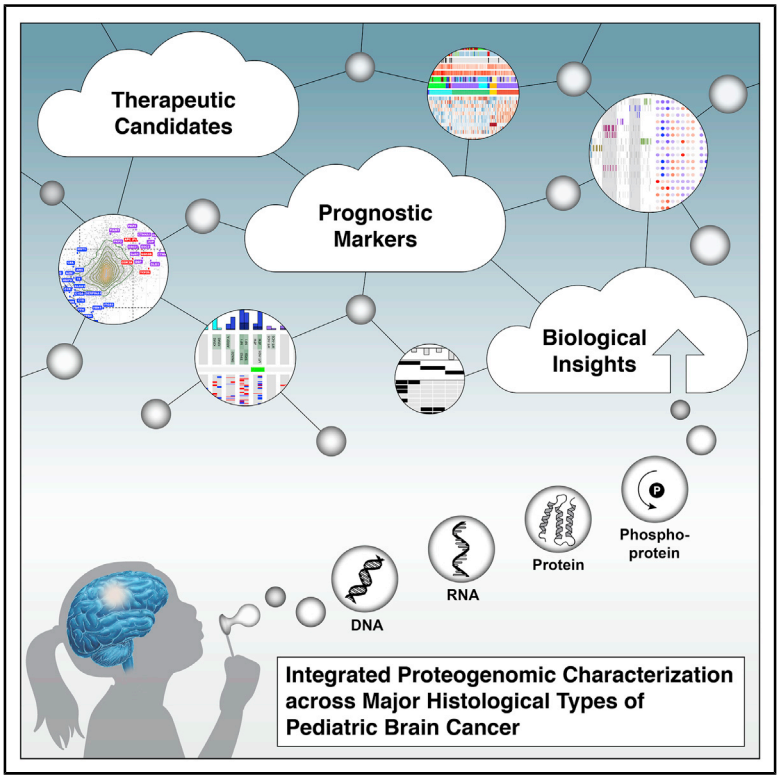
Abstract: We report a comprehensive proteogenomics analysis, including whole-genome sequencing, RNA sequencing, and proteomics and phosphoproteomics profiling, of 218 tumors across 7 histological types of childhood brain cancer: low-grade glioma (n = 93), ependymoma (32), high-grade glioma (25), medulloblastoma (22), ganglioglioma (18), craniopharyngioma (16), and atypical teratoid rhabdoid tumor (12). Proteomics data identify common biological themes that span histological boundaries, suggesting that treatments used for one histological type may be applied effectively to other tumors sharing similar proteomics features. Immune landscape characterization reveals diverse tumor microenvironments across and within diagnoses. Proteomics data further reveal functional effects of somatic mutations and copy number variations (CNVs) not evident in transcriptomics data. Kinase-substrate association and co-expression network analysis identify important biological mechanisms of tumorigenesis. This is the first large-scale proteogenomics analysis across traditional histological boundaries to uncover foundational pediatric brain tumor biology and inform rational treatment selection.

DOI: https://doi.org/10.1016/j.cell.2020.10.044

Cell

Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer

Graphical Abstract



Highlights

- Proteogenomics characterization of 218 pediatric brain tumor samples of 7 histologies
- Proteomic clusters reveal actionable biological features spanning histological boundaries
- Proteomics reveal downstream effects of DNA alterations not evident in transcriptomics
- Kinase activity analyses provide insights into pathway activities and druggable targets

Authors

Francesca Petralia, Nicole Tignor, Boris Reva, ..., Pei Wang, Children's Brain Tumor Network, Clinical Proteomic Tumor Analysis Consortium

Correspondence

resnick@email.chop.edu (A.C.R.), storm@email.chop.edu (P.B.S.), brood@childrensnational.org (B.R.R.), pei.wang@mssm.edu (P.W.)

In Brief

Integrative proteogenomics analysis of pediatric tumors identifies common underlying biological processes and potential treatments as well as the functional effects of somatic mutations and CNVs driving tumorigenesis.

Petralia et al., 2020, Cell *183*, 1–24
December 23, 2020
https://doi.org/10.1016/j.cell.2020.10.044

CellPress

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

Cell

CellPress
OPEN ACCESS

Resource

Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer

Francesca Petralia,[1,35] Nicole Tignor,[1,35] Boris Reva,[1,35] Mateusz Koptyra,[2,33,35] Shrabanti Chowdhury,[1,35] Dmitry Rykunov,[1,36] Azra Krek,[1,36] Weiping Ma,[1,36] Yuankun Zhu,[2,33,36] Jiayi Ji,[3,4] Anna Calinawan,[1] Jeffrey R. Whiteaker,[5] Antonio Colaprico,[6] Vasileios Stathias,[8] Tatiana Omelchenko,[9] Xiaoyu Song,[3,4] Pichai Raman,[2,10,33] Yiran Guo,[2,33] Miguel A. Brown,[2,33] Richard G. Ivey,[5] John Szpyt,[11] Sanjukta Guha Thakurta,[11] Marina A. Gritsenko,[12] Karl K. Weitz,[12] Gonzalo Lopez,[1] Selim Kalayci,[1] Zeynep H. Gümüş,[1] Seungyeul Yoo,[1] Felipe da Veiga Leprevost,[13] Hui-Yin Chang,[13] Karsten Krug,[14] Lizabeth Katsnelson,[15] Ying Wang,[15] Jacob J. Kennedy,[5] Uliana J. Voytovich,[5] Lei Zhao,[5] Krutika S. Gaonkar,[2,10,33] Brian M. Ennis,[2,33] Bo Zhang,[2,33] Valerie Baubet,[2,33] Lamiya Tauhid,[2,33] Jena V. Lilly,[2,33] Jennifer L. Mason,[2,33] Bailey Farrow,[2,33] Nathan Young,[2,33] Sarah Leary,[5,16,17] Jamie Moon,[12] Vladislav A. Petyuk,[12]

(Author list continued on next page)

[1]Department of Genetics and Genomic Sciences and Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[2]Center for Data-Driven Discovery in Biomedicine, Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
[3]Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[4]Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[5]Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
[6]Department of Public Health Science, University of Miami Miller School of Medicine, Miami, FL 33136, USA
[7]Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL 33136, USA
[8]Department of Pharmacology, Institute for Data Science and Computing, Sylvester Comprehensive Cancer Center, University of Miami, Miami, FL 33146, USA
[9]Cell Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[10]Department of Bioinformatics and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
[11]Thermo Fisher Scientific Center for Multiplexed Proteomics, Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA
[12]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA
[13]Department of Pathology, University of Michigan, Ann Arbor, MI 48109, USA
[14]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02412, USA
[15]Institute for Systems Genetics; Department of Biochemistry and Molecular Pharmacology, NYU Grossman School of Medicine, New York, NY 10016, USA

(Affiliations continued on next page)

SUMMARY

We report a comprehensive proteogenomics analysis, including whole-genome sequencing, RNA sequencing, and proteomics and phosphoproteomics profiling, of 218 tumors across 7 histological types of childhood brain cancer: low-grade glioma (n = 93), ependymoma (32), high-grade glioma (25), medulloblastoma (22), ganglioglioma (18), craniopharyngioma (16), and atypical teratoid rhabdoid tumor (12). Proteomics data identify common biological themes that span histological boundaries, suggesting that treatments used for one histological type may be applied effectively to other tumors sharing similar proteomics features. Immune landscape characterization reveals diverse tumor microenvironments across and within diagnoses. Proteomics data further reveal functional effects of somatic mutations and copy number variations (CNVs) not evident in transcriptomics data. Kinase-substrate association and co-expression network analysis identify important biological mechanisms of tumorigenesis. This is the first large-scale proteogenomics analysis across traditional histological boundaries to uncover foundational pediatric brain tumor biology and inform rational treatment selection.

INTRODUCTION

Pediatric brain tumors are the leading cause of cancer-related deaths in children (Ostrom et al., 2018). Although genomic techniques have begun to illuminate the pathogenesis of many pediatric brain tumors, there are some unique challenges that limit the translation of these findings into new effective therapies. Because pediatric brain tumors have a relatively low mutational burden (Chalmers et al., 2017; Gröbner et al., 2018; Northcott et al., 2017; Parsons et al., 2011; Pugh et al., 2012; Robinson et al.,

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

CellPress
OPEN ACCESS

Cell
Resource

Javad Nazarian,[18,19] Nithin D. Adappa,[20] James N. Palmer,[20] Robert M. Lober,[21] Samuel Rivero-Hinojosa,[18] Liang-Bo Wang,[22,23] Joshua M. Wang,[15] Matilda Broberg,[15] Rosalie K. Chu,[12] Ronald J. Moore,[12] Matthew E. Monroe,[12] Rui Zhao,[12] Richard D. Smith,[12] Jun Zhu,[1] Ana I. Robles,[25] Mehdi Mesri,[25] Emily Boja,[25] Tara Hiltke,[25] Henry Rodriguez,[25] Bing Zhang,[26,27,28] Eric E. Schadt,[1] D.R. Mani,[14] Li Ding,[22,23,24] Antonio Iavarone,[29] Maciej Wiznerowicz,[30,31] Stephan Schürer,[8] Xi S. Chen,[6,7] Allison P. Heath,[2,33] Jo Lynne Rokita,[2,10,33] Alexey I. Nesvizhskii,[13,32] David Fenyö,[15] Karin D. Rodland,[12,34] Tao Liu,[12] Steven P. Gygi,[11,37] Amanda G. Paulovich,[5,37] Adam C. Resnick,[2,33,*] Phillip B. Storm,[2,33,*] Brian R. Rood,[18,*] and Pei Wang[1,38,*] Children's Brain Tumor Network, and Clinical Proteomic Tumor Analysis Consortium

[16]Cancer and Blood Disorders Center, Seattle Children's Hospital, Seattle, WA 98105, USA
[17]Department of Pediatrics, University of Washington, Seattle, WA 98195, USA
[18]Children's National Research Institute, George Washington University School of Medicine, Washington, DC 20010, USA
[19]Department of Oncology, Children's Research Center, University Children's Hospital Zürich, Zürich 8032, Switzerland
[20]Department of Otorhinolaryngology, University of Pennsylvania, Philadelphia, PA 19104, USA
[21]Department of Neurosurgery, Dayton Children's Hospital, Dayton, OH 45404, USA
[22]Department of Medicine, Washington University in St. Louis, St. Louis, MO 631110, USA
[23]McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO 63108, USA
[24]Siteman Cancer Center, Washington University in St. Louis, St. Louis, MO 63110, USA
[25]Office of Cancer Clinical Proteomics Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA
[26]Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, TX 77030, USA
[27]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
[28]Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA
[29]Institute for Cancer Genetics, Department of Neurology, Department of Pathology and Cell Biology, Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, NY 10032, USA
[30]Poznan University of Medical Sciences, 61-701 Poznań, Poland
[31]International Institute for Molecular Oncology, 61-203 Poznań, Poland
[32]Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA
[33]Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
[34]Department of Cell, Developmental, and Cancer Biology, Oregon Health & Science University, Portland, OR 97221, USA
[35]These authors contributed equally
[36]These authors contributed equally
[37]Senior author
[38]Lead Contact
*Correspondence: resnick@email.chop.edu (A.C.R.), storm@email.chop.edu (P.B.S.), brood@childrensnational.org (B.R.R.), pei.wang@mssm.edu (P.W.)
https://doi.org/10.1016/j.cell.2020.10.044

2012), the majority of pediatric brain tumors defy treatment approaches that exploit targetable genomic events. In addition, many pediatric brain tumors are characterized by aberrant epigenetic landscapes, but so far there is no effective way to specifically target these key programmatic changes (Capper et al., 2018). RNA profiling has identified subgroups within histological diagnoses and highlighted pathways thought to be active in these groups, but targeting these pathways has largely been unsuccessful. A potential explanation for this lack of translation is that these mechanisms reside many regulatory layers away from the primary functional element of the cell, the protein (Rivero-Hinojosa et al., 2018).

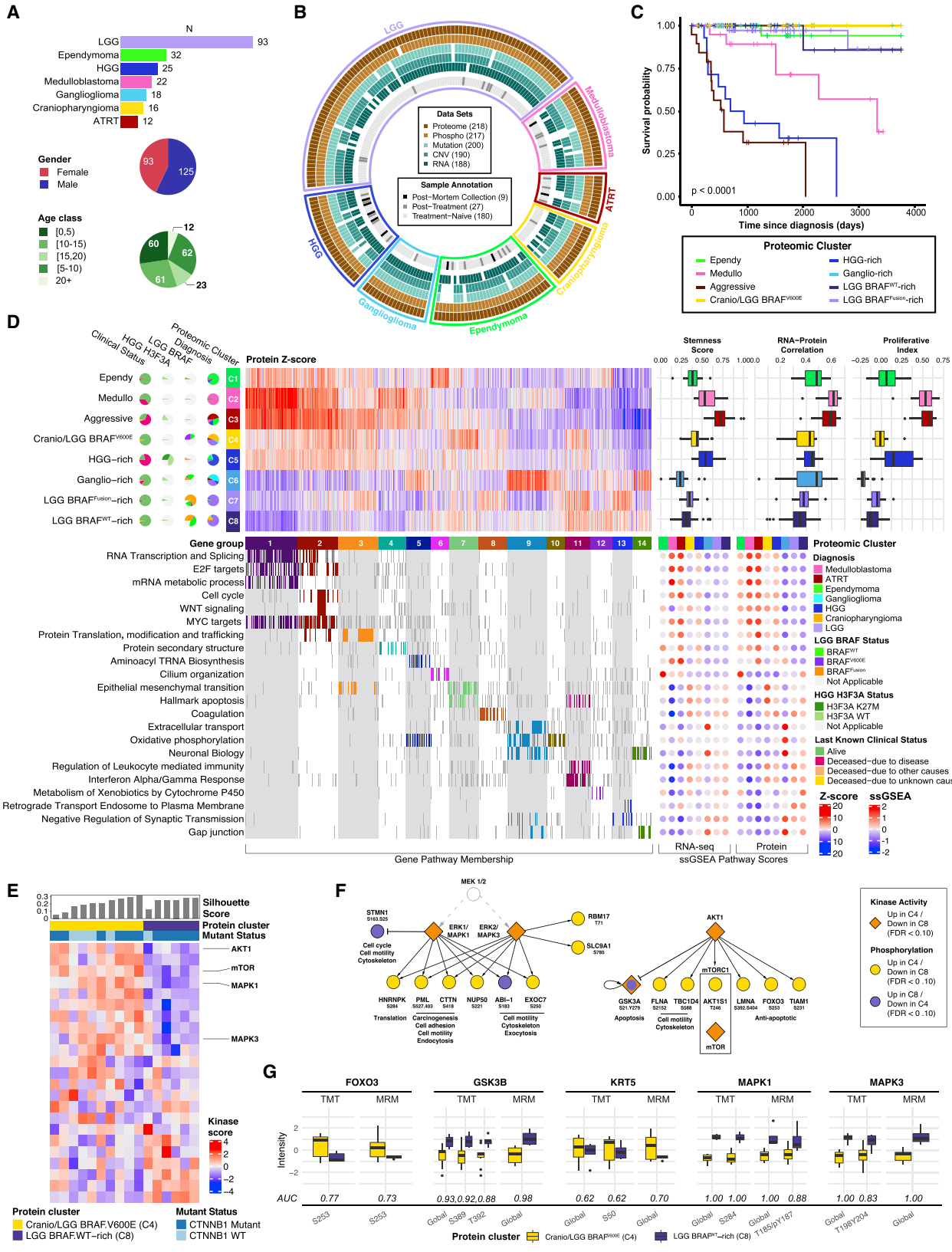
In recent years, quantitative mass spectrometry and bioinformatics analyses have matured, resulting in the ability to add a quantitative proteomics facet to a primarily genomics-based biological understanding of diseases (Clark et al., 2019; Dou et al., 2020; Gillette et al., 2020; Mertins et al., 2016; Rivero-Hinojosa et al., 2018; Zhang et al., 2014, 2016). These efforts have shown a distinct uncoupling of RNA transcript abundance from protein abundance, particularly in cancer. This fact alone could account for a significant disconnect between genome-based biological discovery and clinical validation. Analysis of these integrated proteogenomics datasets has the potential to aid with identification of new therapeutic avenues.

Another challenge of translating new molecular findings into therapeutic innovations is that the subdivision of traditional histology-based entities into molecular subgroups fragments patient populations into ever smaller groups, creating existential challenges for clinical trial design. This is especially true for rare cancers such as pediatric brain tumors. One salient feature of proteomics is the ability to discern biology closer to cellular intent by virtue of its focus on the main functional moiety of the cell, the protein. Because disparate upstream genomic events can result in similar downstream pathway changes, by scrutinizing these resultant events, proteomics can identify common biology across histological and molecular boundaries.

In an attempt to incorporate proteomics into a biological understanding of pediatric brain tumors, we undertook the first large-scale comprehensive proteogenomics analysis inclusive of the genomics, transcriptomics, and global and phosphoproteomics of a large cohort of 218 tumor samples representing 7 distinct histological diagnoses, including low-grade glioma (LGG), ependymoma (EP), high-grade glioma (HGG), medulloblastoma (MB), ganglioglioma, craniopharyngioma (CP), and atypical teratoid rhabdoid tumor (ATRT). Unsupervised clustering based on the proteome revealed surprising alignments between subsets of tumor diagnoses regarded previously as biologically distinct and

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

CellPress
OPEN ACCESS



(legend on next page)

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

led to a number of insights described here. We seek to demonstrate that incorporation of the proteomic and phosphoproteomic dimensions into this large-scale multi-omics study leads to functional insights that will help drive translational efforts.

RESULTS

Proteogenomics Analyses of Pediatric Brain Tumor Specimens

For 218 fresh-frozen tumor samples from 199 patients representing 7 histological types of pediatric brain tumors, we performed whole-genome sequencing (WGS), RNA sequencing (RNA-seq), and quantitative proteomics and phosphoproteomics profiling. All samples were sourced from Children's Hospital of Philadelphia. Figure 1A illustrates the sample distribution across 7 histological types: LGG (n = 93), EP (32), HGG (25), MB (22), ganglioglioma (18), CP (16), and ATRT (12).

For proteomics and phosphoproteomics quantitation, all 218 tissue samples were analyzed by liquid chromatography and triple mass spectrometry with tandem mass tag (TMT) isobaric labeling. The number of proteins and phosphosites measured per sample ranged from 4,661 to 5,731 (median, 5,122) and 2,155 to 3,415 (median, 2,714), respectively. In total, we identified and quantified 8,802 proteins and 18,235 phosphosites. Among them, 6,429 proteins and 4,548 phosphosites were observed in more than 50% of the samples of at least one histological diagnosis and were considered in the downstream analysis. In addition, 440 phosphosites from ischemia-induced proteins (Mertins et al., 2014) were excluded to avoid any artificial effect induced by variations in sample collection.

WGS and RNA-seq were also performed for most samples. After quality filtering, somatic mutation, DNA copy number alterations and RNA-seq-based gene expression data were derived for 200, 190, and 188 tumor samples, respectively (Figure 1B; STAR Methods). All processed proteogenomics datasets can be queried, visualized, and downloaded from http://pbt.cptac-data-view.org/.

Proteogenomics Clustering of Pediatric Brain Tumors

Consensus clustering based on global proteomics data identified eight clusters (Figure S1A) with distinct survival outcomes (Figure 1C), stemness scores, proliferation indices, and pathway activities (Figure 1D; Table S1). We termed the eight clusters *Ependy*, *Medullo*, *Aggressive*, *Cranio/LGG-BRAF^{V600E}*, *HGG-rich*, *Ganglio-rich*, *LGG BRAF^{WT}-rich*, and *LGG BRAF^{Fusion}-rich*.

Although some clusters coincided with histological diagnoses, such as *Medullo*, other clusters contained a mixture of different diagnoses (Figures S1B). First, the *Cranio/LGG BRAF^{V600E}* cluster (C4) aligned a subset of CP tumors with LGG tumors harboring *BRAF^{V600E}* mutations, whereas the rest of CPs were aligned with the *LGG BRAF^{WT}-rich* cluster (C8) (Figures 1D and S1A). This segregation of the CP samples into two distinct clusters was also supported by parallel consensus clustering analysis based on phosphorylation data (Figures S1A). Division of CP samples, however, was not detected based on RNA data (Figure S1A), and the sample-wise correlation between proteomics and RNA-seq profiles was rather low for samples in these two proteomic clusters (Figure 1D).

Although *CTNNB1* mutation is an important oncogenic factor for pediatric CP (Campanini et al., 2010), the proteomic clusters, C4 and C8, did not distinguish *CTNNB1* mutation status (Figure 1E). Instead, they more closely resembled the patterns induced by *BRAF^{V600E}* in LGG (Figures 1D, and 5A). *BRAF^{V600E}* mutation, an oncogenic event for some adult CPs (Brastianos et al., 2014), has not been detected previously in pediatric individuals with CP. Our findings suggest that a subset of pediatric CP tumors, despite their lack of *BRAF^{V600E}* mutations, showed similar proteomics changes as those in *BRAF^{V600E}* LGG tumors. This motivates the hypothesis that some pediatric CPs might benefit from MEK inhibitor (MEKi)-based treatment, a strategy that has been used for *BRAF^{V600E}* LGG tumors (Fangusaro et al., 2019) and has shown preclinical promise in adult CPs (Apps et al., 2018). Indeed, a group of genes suggested to be downregulated by MEKis (Pratilas et al., 2009) has been found to be upregulated in CP samples from C4 (Figure S1C). Furthermore, downstream proteins/substrates of MEK/ERK kinases, including ERK1/2, were upregulated in these samples (Figures 1E, 1F, and S1C), a known consequence of *BRAF^{V600E}* mutation.

In addition, central members of the AKT/mTOR pathway also showed higher kinase activity in C4 compared with C8 CP samples (Figures 1E, 1F, and S1C), consistent with the contrast between *BRAF^{V600E}* and *BRAF^{WT}* LGG tumors (Figure 5B). The AKT

Figure 1. Proteomics-Based Clustering of Pediatric Brain Tumors

(A) Summary of the pediatric brain tumor cohort.

(B) Presence of omics datasets for each of the 218 tumor samples. For each sample, the clinical status at sample collection (i.e., post-mortem, post-treatment, or treatment-naive) is also reported.

(C) Kaplan-Meier curves for overall survival (OS) of patients stratified by proteomic cluster.

(D) Proteomic clusters and differentially expressed proteins allocated to 14 gene groups (top heatmap). Each row represents a proteomic cluster, and each column represents a protein. Red/blue denote up/downregulation patterns of different proteins in a cluster. Distributions of diagnoses, clinical outcomes, and mutation status among the 8 clusters (top left pie plots) and gene members of key pathways enriched in each gene group (bottom heatmap) are shown. For each pathway, the averaged sample-specific gene set enrichment analysis (ssGSEA) score in each proteomic cluster based on global proteomics (protein) and RNA-seq data are illustrated on the right.

(E) Heatmap of kinase activity scores for CP tumors (n = 16). Silhouette scores (top) measure the cohesiveness of tumors classified as C4 and C8 based on kinase activity score. Kinases involved in AKT1 or ERK1/2 signaling are highlighted in the heatmap.

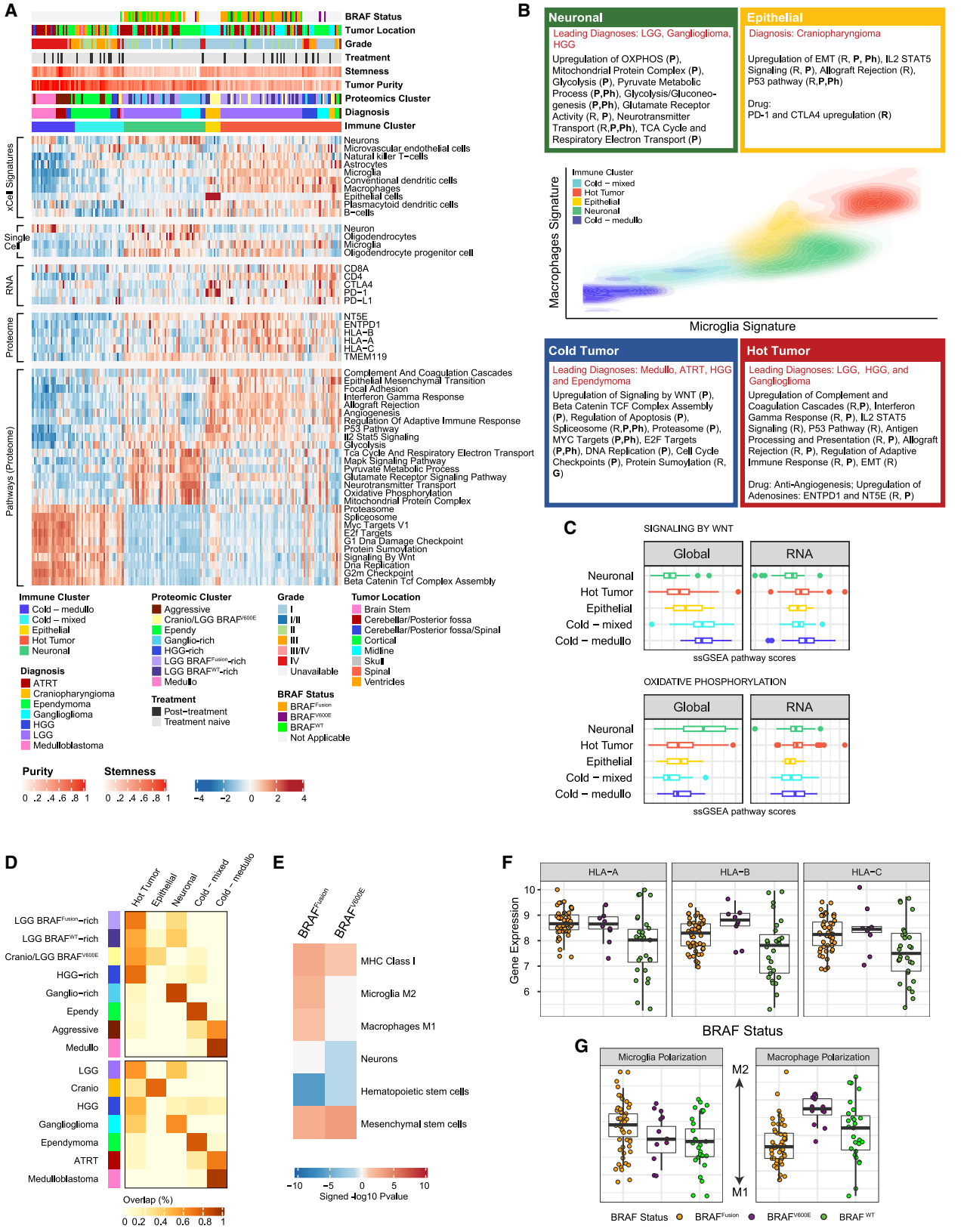
(F) Diagram illustrating differences between C4 and C8 CP tumors in terms of phosphorylation abundance and kinase activity for AKT and ERK1/2 signaling members.

(G) MRM measurements validated different activities of proteins and phosphoproteins between C4 and C8 CPs. The numbers annotated under each pair of boxplots correspond to the AUC (area under the curve) for classifying the two groups of CP using the corresponding protein/phosphosite measurement.

See also Figure S1.

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

CellPress
OPEN ACCESS



(legend on next page)

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

pathway has been implicated as a resistance pathway emerging after RAF/ERK inhibition in BRAF-driven tumors (Jain et al., 2017). Preclinical studies have demonstrated the value of coordinated inhibition of MEK and mTOR, the primary AKT effector, in LGG (Jain et al., 2017). Our findings further suggest potential application of this rationale for some individuals with CP.

Upregulation of key MEK/ERK/AKT kinases in C4 compared with C8 is only visible based on kinase activity assessment using phosphoproteomics data but is not reflected in RNA/protein abundance, suggesting the important complementary role of phosphoproteomics data (Figure S1C).

To validate TMT measurements of proteins and phosphosites of interest, targeted mass spectrometry experiments of a customized protein/phosphoprotein marker panel were applied to the same set of tumor samples following immuno-multiple reaction monitoring (MRM) experiment protocols (Whiteaker et al., 2018). The MRM measurements of key players in the MEK/ERK pathways confirm the substantial differences in C4 and C8 CP (Figure 1G). Moreover, these MRM assays can accurately classify the two subtypes of CP, as reflected by their high area under the curve (AUC) values (Figure 1G), suggesting feasibility of classifying these subtypes in clinical practice by using MRM-based assays.

Another proteomic cluster containing a mixture of diagnoses is the *Aggressive* cluster, characterized by poor survival outcomes (Figure 1C). EPs in the *Aggressive* cluster were more similar to tumors within the cluster, regardless of histology, than to the other EPs in the *Ependy* cluster (Figure S1A). Specifically, members belonging to the evolutionarily conserved multifunctional polymerase-associated factor 1 complex (PAF1C), including PAF1, CDC73, CTR9, LEO1, and RTF1, were found to be upregulated significantly in the *Aggressive* cluster compared with the *Ependy* cluster (Figures S1D and S1E). PAF1C has a vital role in gene regulation and has been implicated in tumorigenesis (Moniaux et al., 2006; Tomson and Arndt, 2013). PAF1C regulates a variety of factors involved in histone covalent modifications, transcription, and mRNA 3′ end processing (Karmakar et al., 2018; Figure S1D). These factors showed upregulation patterns in the *Aggressive* cluster compared with the *Ependy* cluster based on global and phosphoproteomics data (Figure S1E). Neither segregation of EPs into different clusters nor consistent upregulation of PAF1C members and downstream players was observed in RNA data (Figures S1A and S1E).

Although the 9 samples from the post-mortem collection (Figure 1B) blended well with other surgically obtained samples in the protein/RNA-based clustering results, they grouped together in one phosphoproteomic cluster (Figure S1A), suggesting caution when studying post-translational modification (PTM) based on post-mortem samples. To avoid any potential artificial effects, the post-mortem samples were not considered in downstream analyses involving phosphoproteomics data.

Immune Infiltration in Pediatric Brain Tumors

We performed cell type deconvolution analysis using xCell (Aran et al., 2017) based on RNA data to infer relative abundance of different cell types in the tumor microenvironment (Figure 2A; Table S2). The inferred proportion of neuronal and microglia cells were further confirmed based on signatures derived from single-cell RNA-seq study of glioblastoma (Figure 2A; Darmanis et al., 2017). Consensus clustering based on inferred cell proportion identified five sets of tumors with distinct immune and stromal features: *Cold-medullo*, *Cold-mixed*, *Neuronal*, *Epithelial*, and *Hot* (Figure 2A). Comparing this with proteomic clusters, we observed lower immune infiltration in more aggressive proteomic clusters such as *Aggressive*, *Medullo*, and *Ependy* and higher immune infiltration in *LGG BRAF^{WT}-rich*, *LGG BRAF^{Fusion}-rich*, and *Cranio/LGG BRAF^{V600E}* (Figures 2D and S2A).

The *Hot* group, containing a mixture of LGG, HGG, and ganglioglioma samples, was characterized by the presence of multiple types of immune cells, including macrophages, microglia, and dendritic cells (Figure 2A). As expected, compared with other tumors, the *Hot* cluster showed upregulation of immune-related pathways, including epithelial-mesenchymal transition (EMT) (Lou et al., 2016; Figures 2A and 2B). Moreover, adenosine producers (e.g., phosphatases ENTPD1 and NT5E), which have been shown to protect against inflammatory oxidative stress, inhibit immune activators and activate immunosuppressing cells (Chisci et al., 2017; Kordaβ et al., 2018), were upregulated based on RNA and protein data in the *Hot* cluster (Figures 2A and 2B), suggesting that adenosine-reducing therapies can be investigated for these tumors (Lakka and Rao, 2008; Leone and Emens, 2018; Perrot et al., 2019).

Neuronal also contained a mixture of LGG, HGG, and ganglioglioma tumors, but was uniquely characterized by upregulation of glutamate receptor signaling and neurotransmitter transport

Figure 2. Immune Infiltration in Pediatric Brain Tumors

(A) Heatmap illustrating cell type compositions and activities of selected individual gene/proteins and pathways across 5 immune clusters. The heatmap in the first section illustrates the immune/stromal signatures from xCell. The heatmap in the second section illustrates signatures of microglia, neurons, and oligodendrocytes derived from single-cell sequencing data from Darmanis et al. (2017). RNA and protein abundance of key immune-related markers and ssGSEA scores based on global proteomics data for biological pathways upregulated in different immune groups are illustrated in the remaining sections.

(B) Contour plot of two-dimensional density based on macrophage (y axis) and microglia scores (x axis) for different immune clusters. For each immune cluster, key upregulated pathways significant at 10% false discovery rate (FDR) are reported based on RNA-seq (R), global proteomics (P), and phosphoproteomics data (Ph) in the annotation boxes. For the Cold-mixed and Cold-medullo clusters, pathways upregulated in both clusters are reported.

(C) Distribution of pathway scores of signaling by WNT and oxidative phosphorylation based on global proteomics data and RNA stratified by immune clusters.

(D) Heatmap showing the comparison between immune clusters (columns) with proteomic clusters and different histologies (rows). Each row sums to one, with different entries showing the proportion of tumors allocated to different immune clusters.

(E) xCell immune/stromal and antigen presentation signatures in *BRAF^{V600E}* or *BRAF^{Fusion}* compared with *BRAF^{WT}* in LGG.

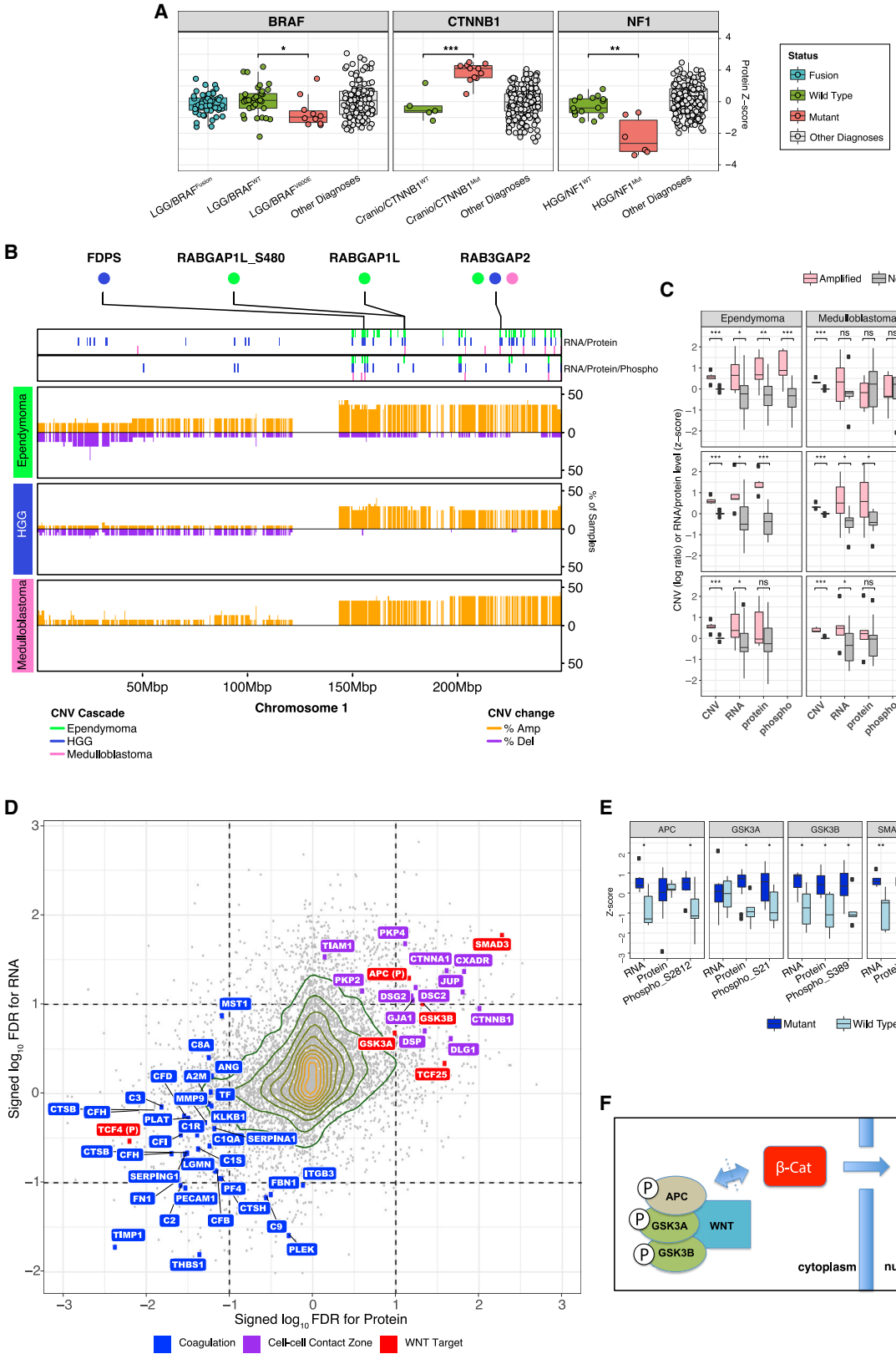
(F) Distribution of RNA levels of *HLA-A*, *HLA-B*, and *HLA-C* in LGG tumors with different *BRAF* statuses.

(G) Distribution of macrophage and microglia polarization (M2-M1) in LGG tumors with different *BRAF* statuses.

See also Figure S2.

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

CellPress
OPEN ACCESS



(legend on next page)

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

pathways involved in neuron communication and activation of cell growth (Pereira et al., 2017; Stepulak et al., 2014; Figures 2A and 2B; Table S2). This observation may support a glutamate/glutamate-receptor mediated mechanism of glioma progression in these samples. Recent studies demonstrate that glutamatergic synapses exist between neurons and glioma cells in pediatric (Venkatesh et al., 2019) and adult (Venkataramani et al., 2019) HGG. Glutamate receptors can activate Ca^{2+}/calmodulin-dependent protein kinase II (CaMK2A/B/G/D), which engages phosphatidylinositol 3-kinase (PI3K; *PIK3CA*) and signals to RAS through centaurin-α1 (*ADAP1*) (Hayashi et al., 2006). Consistently, we found that *GRIA1*, *CAMK2A/B/G/D*, *PIK3CA*, and *ADAP1* showed significant upregulation in *Neuronal* (Table S2), further suggesting an active role of glutamate signaling in the *Neuronal* group. In addition, high levels of glutamate can promote immune evasion mechanisms (Cai et al., 2018); indeed, we observed decreased gene expression of *CD4*, *CD8A*, and macrophage-related genes in the *Neuronal* cluster compared with the *Hot* cluster (Table S2). At the same time, the *Neuronal* cluster was characterized by upregulation of pathways of energy metabolism such as oxidative phosphorylation (OXPHOS), mitochondrial protein complex, and glycolysis solely based on proteomics data (Figures 2A–2C and S2C; Table S2). It has been reported that glutamine blockade induces divergent metabolic programs to overcome tumor immune evasion and that glutamine antagonism could serve as a "metabolic checkpoint" for tumor immunotherapy (Leone et al., 2019), which might benefit tumors like the ones in the *Neuronal* cluster.

LGG tumors, which were split into the *Neuronal* and *Hot* clusters, showed substantial tumor microenvironment heterogeneity (Figure 2A). Interestingly, $BRAF^{V600E}$ and $BRAF^{Fusion}$ events, important oncogenic drivers of LGG tumors, showed significant association with multiple immune signatures. In particular, APM class I genes were upregulated in $BRAF^{Fusion}$ and $BRAF^{V600E}$ tumors compared with the wild-type (Figures 2E and 2F; Table S2). More careful investigation of pro-inflammatory (M1) and pro-regenerative (M2) macrophage and microglia signatures (Figure S2B; Table S2) based on markers specific to these cell types (Dello Russo et al., 2017; Fumagalli et al., 2018; Krasemann et al., 2017) further suggests that M1 macrophages and M2 microglia were upregulated in $BRAF^{Fusion}$ compared with the wild-type (Figure 2E). The significant difference between microglia and macrophage polarization across BRAF statuses is further illustrated in Figure 2G; $BRAF^{Fusion}$ promoted more M2 microglia, whereas $BRAF^{V600E}$ promoted more M2 macrophages. This observation is in concordance with the balance between macrophage and microglia polarization reported for adult glioblastoma (Darmanis et al., 2017).

The *Epithelial* cluster, containing, as expected, only CP tumors, which originate from odontogenic epithelium, was characterized by upregulation of EMT, immune-related pathways, as well as CTLA4 and PD-1 molecules (Figures 2A and 2B; Table S2). Therefore, CP could potentially benefit from immune checkpoint therapy, as reported previously (Coy et al., 2018).

Finally, *Cold-medullo* and *Cold-mixed* exhibited upregulation of signaling by WNT, β-catenin TCF complex assembly, regulation of apoptosis, and proteasome. This is consistent with recent reports that tumors with active WNT signaling are characterized by lower levels of immune infiltration (Luke et al., 2019). Again, these patterns of upregulation were observed in *Cold-medullo* and *Cold-mixed* clusters based on proteomics and phosphoproteomics data but not RNA data (Figures 2A–2C and S2C).

Integrative Proteogenomics Analyses Reveal Functional Consequences of Mutation and Copy Number Variation (CNV)

Although pediatric tumors usually have fewer genetic alterations compared with adult tumors (Gröbner et al., 2018), a few recurrent DNA alterations were observed in this cohort (Figure S3A). We first evaluated the effect of the few somatic mutations on the corresponding RNA/protein levels. LGG tumors with $BRAF^{V600E}$ mutation had significantly downregulated BRAF protein abundance compared with $BRAF^{WT}$ LGG tumors (Figure 3A), whereas the reduction was not significant at the transcript level (Figure S3B). *CTNNB1* mutation resulted in elevated protein/RNA levels among CP samples, whereas *NF1* mutation resulted in downregulation of cognate proteins and transcripts in HGG (Figures 3A and S3B). *SMARCB1* RNA/protein were significantly downregulated in ATRT samples compared with other diagnoses as expected, and the downregulation was the result of different types of DNA alterations, including mutation, deletion, and copy-neutral loss of heterozygosity (LOH) (Figure S3C).

In terms of genomic instability, MB, HGG, and EP tumors showed relatively higher genomic instability (Figure S3A). By

Figure 3. Effect of Genomic Alterations on Transcriptomic, Proteomic, and Phosphoproteomic Abundance

(A) Distribution of protein abundance of BRAF, CTNNB1, and NF1 across tumor samples stratified by different mutation status and diagnoses. *$p < 0.1$, **$p < 0.01$, and ***$p < 0.001$.

(B) DNA copy number amplification/deletion frequencies along chromosome 1 among EP, HGG, and MB samples. Genes with detected CNV-RNA/protein or CNV-RNA/protein/phospho cascade events are labeled as vertical bars in the top track.

(C) Distribution of DNA copy number (log ratio), RNA, and protein abundance of RABGAP1L, RAB3GAP2, and FDPS, stratified by their amplification statuses in EP, MB, and HGG tumors. *$p < 0.1$, **$p < 0.01$, and ***$p < 0.001$. ns, not significant (p value > 0.1).

(D) Illustration of the effect of *CTNNB1* mutation on RNA and protein abundance in CP samples. The x axis (y axis) represents signed −log10 FDR for testing the association between protein abundances (RNAs) and *CTNNB1* mutation. The cell-cell contact zone (coagulation) pathway is enriched in the set of proteins upregulated (downregulated) in *CTNNB1* mutant samples. A few members of the WNT signaling pathway whose protein or phosphosites are associated with *CTNNB1* mutation are highlighted in red. Phosphosites are annotated with (P) in their gene symbols.

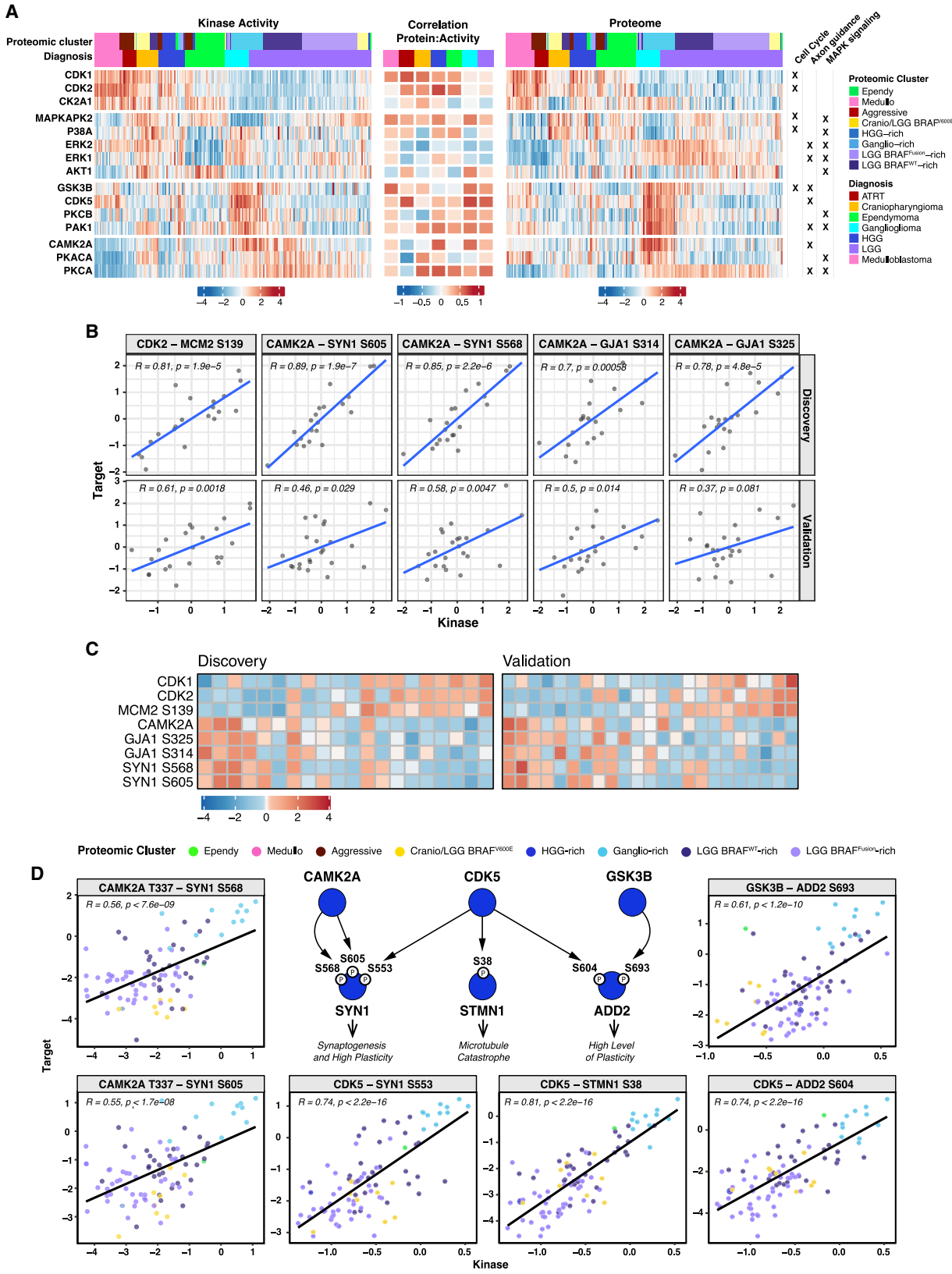
(E) Distribution of protein and phosphosite abundance among *CTNNB1* mutant and *CTNNB1* wild-type CP tumors for known key members of the WNT signaling pathway interacting with β-catenin and transcription factors regulated by *CTNNB1*. Symbols *, **, and *** correspond to an FDR of less than 0.1, 0.01 and 0.001, respectively. ns, FDR > 0.1.

(F) Illustration of the regulatory role of β-catenin.

See also Figure S3.

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

CellPress
OPEN ACCESS



(legend on next page)

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

integrating copy number, RNA, and proteomics data, we detected 1,541 genes whose transcript and protein abundance were influenced simultaneously by their own CNVs in one or more diagnosis, referred to as CNV-RNA/protein *cis*-cascade events (Figure S3D; Table S3). In addition, for 515 of these 1,541 genes, we detected significant dependence between their phosphosite abundance and CNV in one or more diagnoses, referred to as CNV-RNA/protein/phospho *cis*-cascade events (Figure S3D; Table S3). These lists of *cis*-cascade events facilitate identification of important players in frequently amplified/deleted genome regions. One example is *RABGAP1L* (*1q25*), an EP CNV-RNA/protein/phospho *cis*-cascade gene (Figures 3B and 3C), whose amplification is associated with GTPase activation and RAB-GTPase binding (Itoh et al., 2006) and has been reported to be an independent predictor of tumor progression in EP (Kilday et al., 2012). Another member from the *RAB GTPase* gene family, *RAB3GAP2* (*1q41*), which has a key role in neurodevelopment (Ng and Tang, 2008), was identified to be a CNV-RNA/Protein cis-cascade gene for EP, MB, and HGG tumors. Our analysis further pinpoints an important player in maintenance of glioblastoma stemness, *FDPS*, proximate to *RABGAP1L*, as a CNV-RNA/protein *cis*-cascade gene for HGG (Abate et al., 2017; Kim et al., 2018a).

Although recurrent amplification of *RABGAP1L*, *RAB3GAP2*, and *FDPS* was observed in all EP, MB, and HGG tumors, a significant influence of *RABGAP1L* amplification on its protein/phosphoprotein was only observed in EP, whereas *FDPS* was found to be a CNV/RNA/protein cascade event only in HGG (Figures 3B and 3C). On the other hand, for MB, only *RAB3GAP2* is identified as a CNV-RNA/protein *cis*-cascade gene. These observations suggest that CNV of the same genomic region could lead to different functional perturbations in different diagnoses.

We then studied the *trans*-regulatory effects of somatic mutations and CNVs on proteins and phosphoproteins within each diagnosis. Besides *BRAF* mutation/fusion in LGG (discussed below), the only other profound *trans*-regulatory effects were detected between mutation of *CTNNB1*, which codes β-catenin, and many proteins and phosphosites in CP (Figure 3D; Table S3). β-Catenin is crucial for two developmental processes: establishment and maintenance of cell-type-specific cell-to-cell adhesion and regulation of target gene expression via the WNT signaling pathway (Gao et al., 2017). As expected, *CTNNB1* mutation, which boosted β-catenin abundance, was found to be associated with upregulation of proteins/phosphosites related to cell-to-cell adhesion as well as upregulation of members of the WNT signaling pathway, such as APC, GSK3A, and GSK3B (Figures 3D and 3E). Specifically, although phosphosite abundance of APC at *Ser 2812* showed significant elevation in *CTNNB1* mutation cases, this upregulation was not observed based on protein abundance of *APC* (Figure 3E). It is well known that WNT signaling results in liberation of β-catenin and its translocation to the nucleus, where it binds to transcription factor (TCF) complexes to activate transcription (Figure 3F). In our data, we observed significant association between RNA and protein/phosphosite abundance of *TCF4*, *TCF25*, and *CTNNB1* mutation in CP. The interaction of β-catenin with *TCF4* has been proposed as a target for development of anticancer drugs in other tumor types (Fasolini et al., 2003). However, we observed that RNA expression and phosphosite abundance of *TCF4* were significantly lower in the *CTNNB1*-mutated group. RNA and proteomic abundance of *TCF25*, another transcription factor that may play a role in cell death control (Cai et al., 2006), were upregulated in *CTNNB1*-mutated CP. These results suggest that, among this group of CP, downstream effects of mutation in *CTNNB1* could be mediated by *TCF25*.

Phosphoproteomics Analysis of Kinase Activity

Because of the tremendous appeal of kinases as drug targets, it is of great importance to characterize the common and differential kinase activations within and across histologies. CDK1 and CDK2, essential cyclin-dependent kinases promoting the G2-M transition and regulating G1 progression and the G1-S transition (Santamaría et al., 2007), were elevated in more proliferative tumors, including ATRT, MB, HGG, and EP based on global abundance and kinase activity (Figures 4A and S4B; Table S4), the latter derived from the abundance of phosphorylated substrates (STAR Methods). Activation of CDK1 and CDK2 in more proliferative tumors was also confirmed by higher correlation between their global abundance and kinase activity scores (Figure 4A). To further characterize the dependence of individual substrates on kinases, we constructed diagnosis-specific kinase-substrate networks leveraging an experimentally validated kinase-substrate regulation database (Hornbeck et al., 2015; STAR Methods). Some kinase-substrate associations of CDK1/2 were shared across different diagnoses (Figure S4A). For example, the association between CDK2 and MCM2 at Ser 139 was detected in ATRT/MB, EP, HGG, and LGG. On the other hand, some diagnosis-specific associations were detected, such as CDK2 and NPM1 at Ser 70 in HGG and LGG, and CDK2 and TERF2IP at Ser 203 in EP (Figure S4A). MCM2, NPM1, and TERF2IP (*RAP1*) have important roles in cell proliferation (Box et al., 2016; Fei and Xu, 2018; Schmitt and Stork, 2001), implying diverse mechanisms used by CDK2 to influence cell proliferation in various diagnoses.

Figure 4. Phosphoproteomics Analysis of Kinase Activity

(A) Heatmaps showing the global abundance (right panel) and the kinase activity score (left panel) of selected kinases across different histologies. For each kinase, the Pearson's correlation between its global abundance and kinase activity within each histology is shown in the center panel.

(B) Scatterplot showing the global abundance of a particular kinase (x axis) versus the phospho-abundance of the targeted substrates (y axis). The first row is based on data from the discovery cohort, and the second row displays data based on the validation cohort.

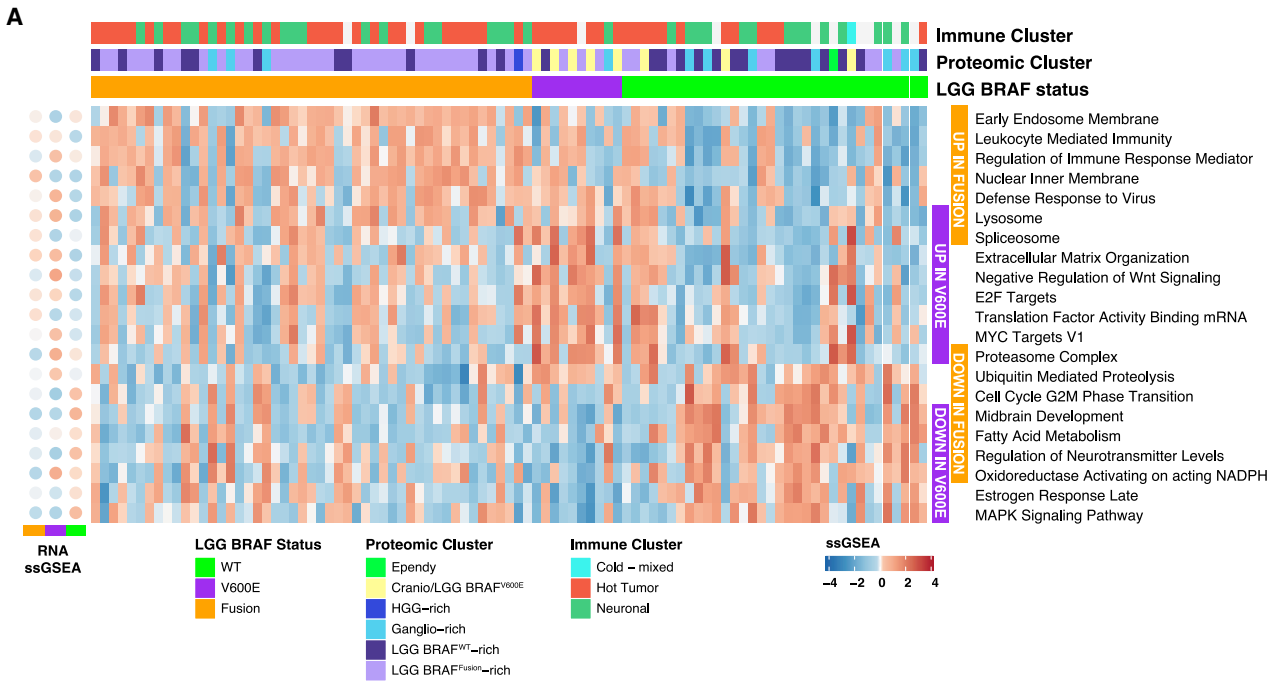
(C) Heatmap showing global proteomic abundance of CDK1, CDK2, and CAMK2A as well as phosphorylation abundance of MCM2 Ser 139, GJA1 Ser 325, GJA1 Ser 314, SYN1 Ser 568, and SYN1 Ser 605 among HGG in the discovery and validation cohorts.

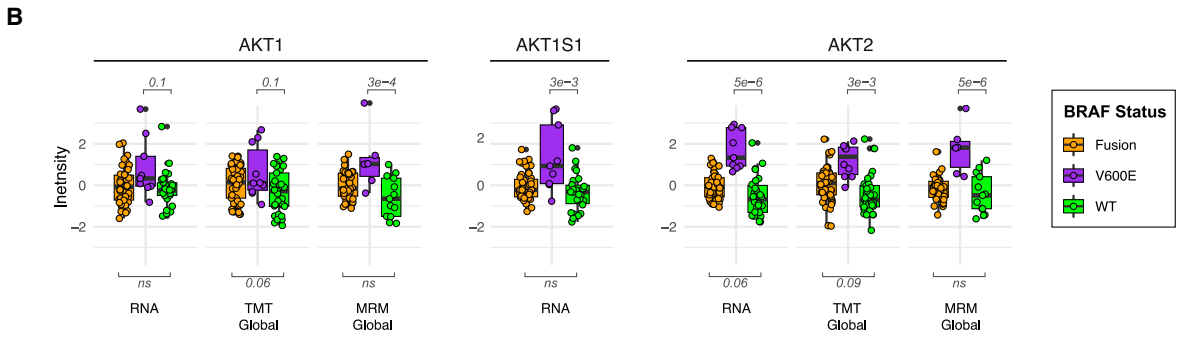
(D) Diagram showing kinase-substrate associations involved in CNS development in LGG (top center panel). Scatterplots show the association between the global (or phospho) abundance of each kinase (x axis) and the phospho-abundance of the corresponding substrate (y axis).

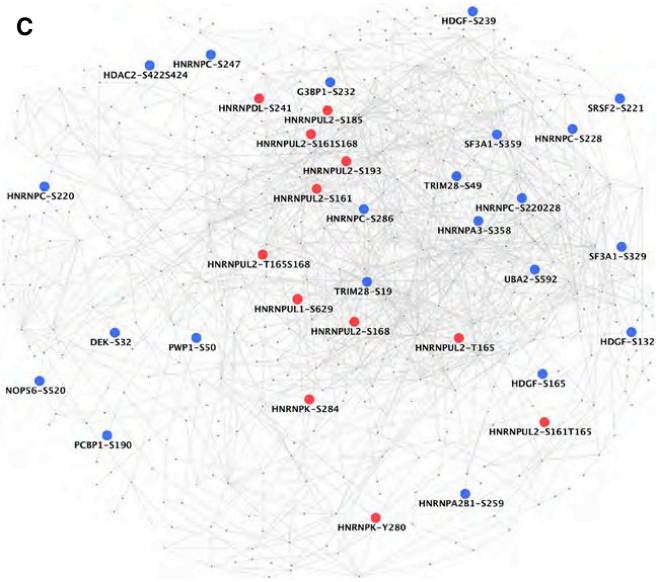
See also Figure S4.

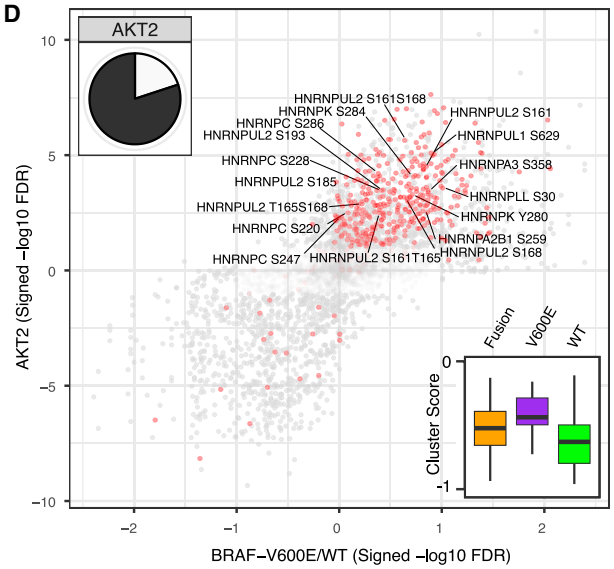
Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

CellPress
OPEN ACCESS









(legend on next page)

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

Another important kinase is CaMK2A (Ca^{2+}/calmodulin-dependent protein kinase II alpha), which is directly involved in metastatic invasion of glioma cells (Chen et al., 2011; Cuddapah and Sontheimer, 2010; Shin et al., 2019). Although CaMK2A was most abundant in ganglioglioma, a higher correlation between its kinase activity score and protein abundance was observed in HGG (Figure 4A; Table S4). The inferred kinase-substrate network of HGG further highlights the association between CaMK2A and GJA1 (connexin 43) at Ser 325 and Ser 314 (Figures 4B and S4A). Phosphorylation of connexin 43 at Ser 325 and Ser 314 promotes gap junction assembly between glioma and astrocytes (Cooper and Lampe, 2002) and drives cancer cell migration as well as glioma invasion (Behrens et al., 2010; Hong et al., 2015). Thus, our data suggest a potential role of CaMK2A in glioma invasion. Moreover, in HGG, CaMK2A protein abundance was found to be associated with SYN1 Ser 568 and SYN1 Ser 605 (Figures 4B and S4A); the latter increases synaptic transmission and regulates synaptic vesicle dynamics (Magupalli et al., 2013). This further aligns with a relevant role of CaMK2A in glioma invasion because glioma cells form functionally active synapses with neurons, and neural activity mediated by neuron-to-glioma synapses drives glioma invasion and growth (Venkataramani et al., 2019; Venkatesh et al., 2019). Interestingly, activation of CDK1/2 and CaMK2A, reflected by their elevated protein abundance, tended to be exclusive of each other, suggesting the existence of two different signaling mechanisms among HGG tumors (Figure 4C).

To further confirm the kinase activity of CDK2 and CAMK2A in HGG, we carried out independent TMT proteomics and phosphoproteomics experiments in an independent cohort of 23 pediatric and young adult HGGs (STAR Methods) and validated the kinase-phosphosite associations between the aforementioned pairs (Figure 4B). The negative correlations between CDK1/2 and CAMK2A protein abundance were also confirmed in this validation cohort (Figure 4C), suggesting two different signaling mechanisms among HGG tumors.

Another interesting group of kinases, CDK5 and GSK3B, were upregulated in ganglioglioma and a subset of LGG belonging to the *Ganglio-rich* cluster (Figure 4A). CDK5 and GSK3B have been suggested to be regulators of synapse formation, neurogenesis, and cell proliferation (Cole, 2012; Shah and Lahiri, 2017). The kinase-phospho network revealed interesting associations between CDK5/GSK3B and their substrates in LGG (Figures 4D and S4A), such as the phosphosites of ADD2 (beta-adducin). ADD2 is highly expressed in brain regions associated with high plasticity (e.g., the hippocampus), involved in neuronal morphology, and required for synaptogenesis (Bednarek and Caroni, 2011; Porro et al., 2010). Positive associations between CDK5 and ADD2 at Ser 604 as well as GSK3B and ADD2 at Ser 693 reflected CDK5-dependent priming of GSK3B activity (Farghaian et al., 2011). Moreover, it has been shown that CDK5 regulates recruitment of SYN1 to nascent synapses (Easley-Neal et al., 2013), and phosphorylation of SYN1 by CDK5 at Ser 553 controls the efficiency of neurotransmitter release (Qiao et al., 2014). Thus, the observed positive association between CDK5 and SYN1 Ser 553 in LGG might support an increase in synaptogenesis between glioma cells and neurons in LGG, which, however, needs to be confirmed further by functional studies. This increase also aligns with the association between CaMK2A1 and SYN1 at Ser 605 and Ser 568 (Figures 4D and S4A) because phosphorylation of SYN1 at Ser 605 by CaMKII has been shown to increase synaptic transmission (Magupalli et al., 2013). Furthermore, the positive association between CDK5 and STMN1 at Ser 38 highlights the importance of STMN1-mediated synaptogenesis for pediatric gliomagenesis because stathmin phosphorylation, including Ser 38, is essential for synaptic plasticity and memory and increases synaptic strength by promoting microtubule stability and dendritic transport of the GluA2 subunit of AMPA-type glutamate receptors to the synapse (Uchida et al., 2014). All of these observations are in line with findings showing that gliomas can hijack neuronal development by creating neuron-glioma synapses (Venkataramani et al., 2019; Venkatesh et al., 2019) and link to the immune clustering results: the global abundance of CDK5 and GSK3B was upregulated in the subset of LGGs from the *Neuronal* immune cluster (Figure S4C).

Insights from Proteogenomics Analysis of LGG

To help discern biological insights stemming from the frequent targetable alterations of *BRAF* in LGG, we identified proteins associated with $BRAF^{V600E}$ mutation and $BRAF^{Fusion}$ (Table S5). Compared with $BRAF^{WT}$ tumors, $BRAF^{V600E}$ and $BRAF^{Fusion}$ cases showed common as well as alteration-type specific changes (Figure 5A). Particularly in $BRAF^{V600E}$, we observed significant abundance changes of proteins in the mitogen-activated protein kinase (MAPK; ERK) signaling pathway (Figure 5A) compared with $BRAF^{WT}$ tumors. MAPKs are the terminus of the RAS/RAF/MAPK pathway, whose inhibitors have been used to treat *BRAF*-altered tumors of multiple cancer types, including brain tumors (Schreck et al., 2019). For instance, MEKi monotherapy recently showed promising results in low-grade pediatric glioma with *BRAF* alterations (Fangusaro et al., 2019). Investigation of an RNA expression-based "MEKi

Figure 5. Insights from Proteogenomics Analysis of LGG

(A) Heatmap illustrating ssGSEA scores of selected pathways differentially expressed between LGG tumors with different *BRAF* statuses based on global proteomics data. The dot plot on the left summarizes ssGSEA pathway scores based on RNA data among samples with different *BRAF* statuses.

(B) Distributions of RNA and TMT protein abundance (TMT global), and MRM protein abundance (MRM global) of AKT1, AKT2, and AKT1S1 in samples with different *BRAF* alteration statuses. FDR levels of two-sample comparisons between $BRAF^{V600E}$/$BRAF^{Fusion}$ and $BRAF^{WT}$ are annotated.

(C) The network topology representing the LGG phosphosite co-expression network module enriched in sites upregulated in $BRAF^{v600E}$ compared with $BRAF^{WT}$ tumors. Phosphosites mapping to genes in the HNRNP family or contained in the MYC target pathway are highlighted in red and blue, respectively.

(D) Scatterplot displaying the association between each phosphosite's abundance with the global abundance of AKT2 (y axis) versus the association with $BRAF^{V600E}$ (x axis). Phosphosites contained in the network module in (C) are highlighted in red. Boxplots illustrate the distribution of the activity scores (ssGSEA) of the network module in (C) based on phosphoproteomics data in samples with different *BRAF* statuses. The pie plot shows the proportion of phosphosites contained in the network module in (C), whose abundances are associated at 5% FDR with the global abundance of AKT2.

See also Figure S5.

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

CellPress
OPEN ACCESS

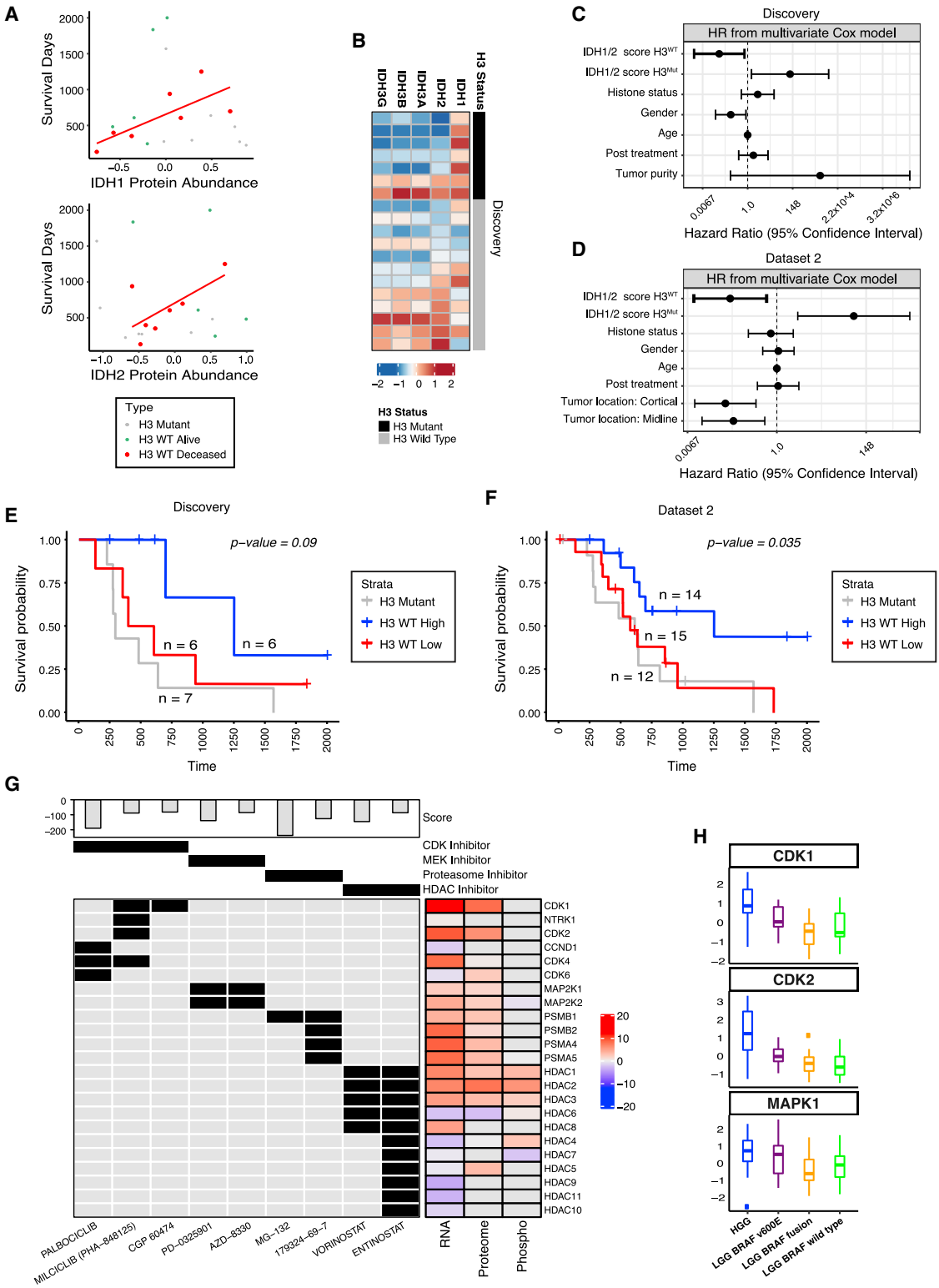


Figure 6. Insights from Proteogenomic analysis of HGG

(A) Scatterplot showing OS of individuals with HGG versus the global protein abundance of IDH1 and IDH2 in the tumors.

(B) Heatmap of the global abundance of IDH proteins in the discovery cohort.

(C and D) 95% CI of hazard ratio coefficients from Cox regression for IDH1/2 scores and other covariates based on the discovery cohort (C) and dataset 2 (D).

(legend continued on next page)

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

signature" (Pratilas et al., 2009) in our data confirmed that genes downstream of MEK kinases are greatly upregulated in *BRAF*^{V600E} compared with *BRAF*^{WT} tumors (Figure S5B), supporting the current use of MEKi therapy for these LGG tumors. Moreover, RNA/protein abundance of the AKT serine/threonine kinases AKT1 and AKT2 as well as RNA of AKT1S1 (Figure 5B; Table S5) showed significant upregulation in *BRAF*^{V600E} tumors. MRM experiments measuring AKT isoforms on the same set of tumors further validated this upregulation (Figure 5B). Indeed, the AKT pathway has been implicated as a resistance pathway emerging after RAF/MAPK inhibition in *BRAF*-driven tumors via upregulation of receptor tyrosine kinases (Jain et al., 2017). Preclinical studies have demonstrated the value of coordinated inhibition of MEK and mTOR, the primary AKT effector, in LGG (Jain et al., 2017). Our findings further strengthen this rationale for upcoming clinical trials.

Next we employed a network-based approach to study the effect of BRAF alterations on the phosphoproteome. Co-expression network analysis resulted in 18 closely connected modules capturing the association across phosphosites (Table S5). Interestingly, two modules (Figures 5C and S5C) are significantly upregulated in *BRAF*^{V600E} and *BRAF*^{Fusion} samples, respectively (Table S5). Module 1 was significantly enriched in phosphosites associated with MYC targets (Figure 5C) and the G2M checkpoint, confirming upregulation of cell cycle-related pathways in individuals with *BRAF*^{V600E} compared with *BRAF*^{WT} LGG (Figure 5A). Moreover, module 1 was significantly enriched in phosphosites regulated by AKT2 (Figure 5D). Specifically, it contained phosphosites of a group of heterogeneous nuclear ribonucleoproteins, including *HNRNPUL1* and *HNRNPUL2 (*Figures 5C and 5D*)*. Active AKT2 has been reported to suppress the interaction between HNRNPU and caspase-9b, causing inhibition of apoptosis (Vu et al., 2013). These observations further support the concept of inhibiting mTOR/AKT in *BRAF*^{V600E} LGG.

On the other hand, module 2 appeared to capture a group of phosphosites perturbed in *BRAF*^{Fusion} but not in *BRAF*^{V600E} cases (Figures S5C–S5E). The top druggable kinase associated with these phosphosites in this module is PDGFRA, which encodes a cell surface tyrosine kinase receptor (Figure S5D). PDGFRA is frequently mutated/amplified in pediatric HGG and has been suggested to serve as a treatment target for pediatric HGG (Koschmann et al., 2016). Our data reveal a strikingly similar upregulation of PDGFRA protein/RNA in *BRAF*^{Fusion} samples as in HGG tumors (Figure S5A), suggesting exploration of *PDGFRA*-targeted treatment in *BRAF*^{Fusion} tumors as well.

Insights from Proteogenomics Analysis of HGG

Isocitrate dehydrogenases (IDHs) are enzymes that catalyze oxidative decarboxylation of isocitrate, producing α-ketoglutarate (KG) and CO_2. Mutations of the IDH1 and IDH2 proteins, which can be effectively targeted by drugs, have been found in ~80% of grade II and III astrocytomas, oligodendrogliomas, and secondary glioblastomas (Bergaggio and Piva, 2019). These mutations, however, are infrequent in pediatric HGG (~11%) (Kim and Liau, 2012), as also observed in our data. On the other hand, recent literature has reported prognostic and/or therapeutic roles of the wild-type IDH genes/proteins in various adult cancers, such as melanoma, glioblastoma, and kidney cancer (Bergaggio and Piva, 2019; Tanaka et al., 2013; Calvert et al., 2017), bringing interest to understanding their roles in pediatric HGG tumors.

We first investigated associations between IDH protein abundances and overall survival (OS) among HGG patients. Because point mutations in histone *H3.3* (*H3F3A*, *H3K27M*) have been reported to lead to a worse prognosis in HGG (Karremann et al., 2018), *H3* status was adjusted when assessing the association between OS and abundance of IDH proteins (STAR Methods). Strikingly, all IDH proteins showed positive association with improved OS among the *H3*^{WT} group (Figures S6A and 6A; Table S6). A parallel analysis based on RNA data detected similar associations between expression of *IDH1/2/3A* with OS (Figure S6B). Consistently, the oxidative phosphorylation pathway, harboring IDH1/2/3, is one of the leading pathways whose upregulation was significantly associated with improved OS among *H3*^{WT} individuals (Figure S6C).

Although all IDH proteins showed positive correlation with OS, no correlation was observed between IDH1 and IDH2/3 protein abundance (Figure 6B). Although this is not surprising, because IDH1 is situated in the cytosol and peroxisomes whereas IDH2/3 are in the mitochondria, it implies potentially complementary information in IDH1 and IDH2/3 for prognostic prediction. Indeed, compensatory functions between IDH1 and IDH2 have been reported in acute myeloid leukemia (Zhang et al., 2019) and colorectal cancer (Koseki et al., 2015). These observations motivated us to evaluate the joint prognostic value of IDH proteins. In addition, because IDH3 proteins are highly correlated with IDH2, and IDH3 proteins are of relatively lower abundance compared with IDH1/2, we decided to focus on IDH1 and IDH2 to avoid collinearity in the analysis. Specifically, by jointly modeling IDH1 and IDH2 proteins in one multivariate Cox regression model, we estimated that, among individuals with *H3*^{WT} HGG, the risk of death has a 23.58-fold increase, with a 95% confidence interval (CI) of [1.42, 384.6] (Figure 6C), when the combined abundance of IDH1 and IDH2 is 50% lower (i.e., a decrease of 1 in the weighted log2 abundance; STAR Methods). The extremely wide CI for the hazard ratio of the IDH1/2 score is a result of the limited sample size in the analysis (n = 19). To further verify this finding, we performed TMT proteomics profiling experiments for an additional 41 pediatric HGG samples, including 23 from an independent cohort, and 18 from the existing study cohort with the remaining

(E and F) Kaplan-Meier curves of OS for HGG *H3*^{Mut} samples (gray), *H3*^{WT} samples with low IDH1/2 abundance (red), and *H3*^{WT} tumors with high IDH1/2 abundance (blue) for the discovery cohort (E) and the validation cohort (F).

(G) Illustration of drug target analysis results. The bottom left heatmap illustrates the targeting genes (rows) of each detected drugs (columns). For each gene, the *Z* score comparing its RNA and proteomic abundances between HGG and LGG is shown in the bottom right heatmap. Mechanism of actions are annotated at the top of the heatmap together with the resulting score from the connectivity map analysis.

(H) Distribution of kinase activity scores of CDK1, CDK2, and MAPK1 among HGG and LGG tumors, with the latter further stratified by BRAF status.

See also Figure S6.

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

CellPress
OPEN ACCESS

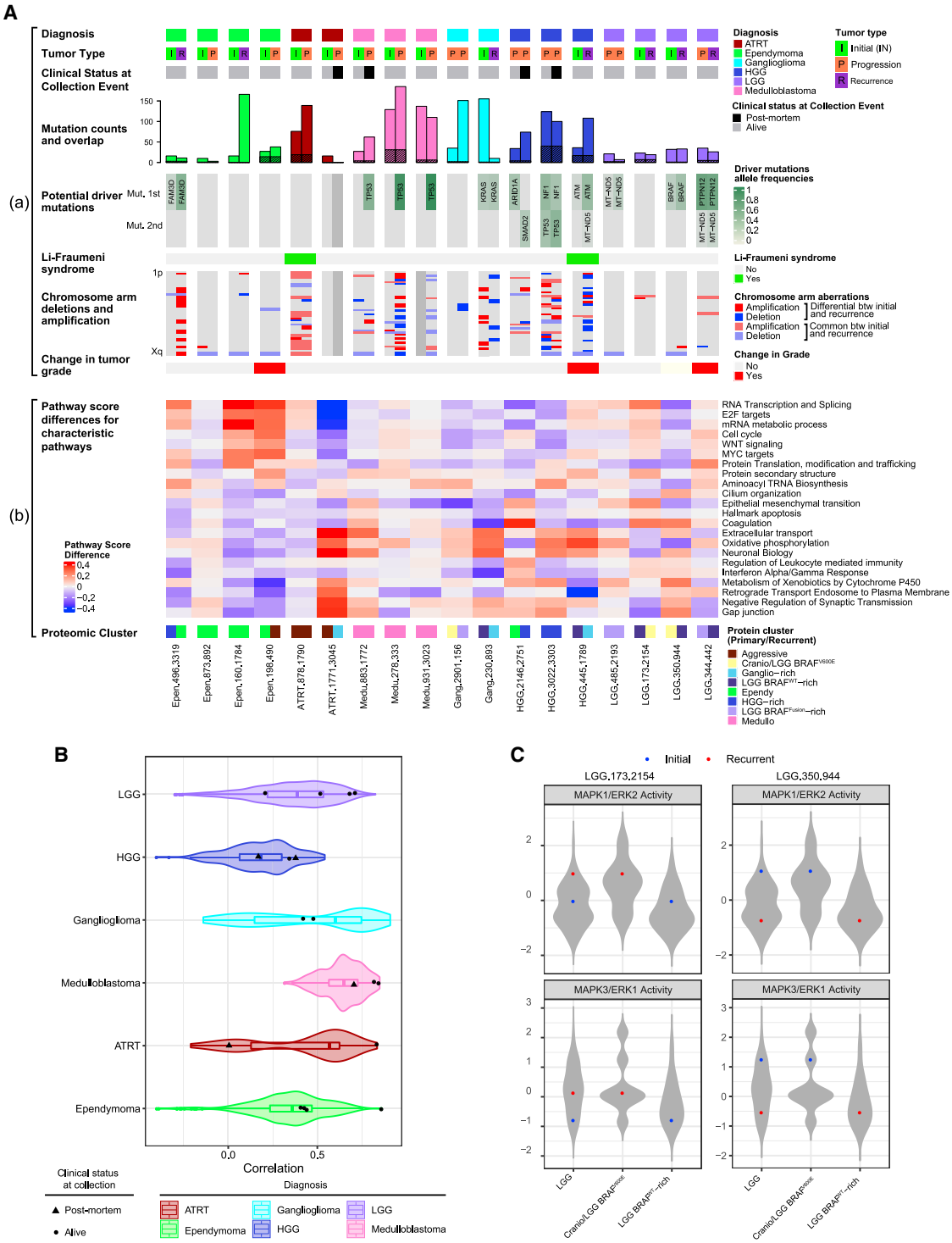


Figure 7. Comparison between IN and Recurrent Tumors

(A) a: clinical properties and genomic characterization of 18 pairs of IN versus RP tumors. The bar plot illustrates the number of non-synonymous mutations in IN and RP tumors, with the number of shared mutations represented by the shaded area. The potential driver mutation track shows the allele frequencies of somatic mutations of known oncogenes and tumor suppressor genes. Chromosome arm aberrations of each sample and the change of tumor grade from IN to RP of each patient are also shown. b: differences in ssGSEA score between RP and IN tumors of key molecular pathways associated with different proteomic clusters. The annotation at the bottom indicates the diagnosis and clinical event IDs of the paired samples for each individual. For example, "Epen.496.3319" refers to a pair of EP tumors with IDs 7316-496 and 7316-3319.

(legend continued on next page)

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

tumor material (STAR Methods). With this second dataset, we confirmed the association between the reduced expression of combined IDH1/2 protein abundance and shorter OS after accounting for additional confounders, such as tumor location (Figures 6D, 6F, and S6E; STAR Methods).

Unlike *H3*^{WT} HGG tumors, IDH1/2 proteins showed an adverse effect on OS among *H3*^{MUT} tumors (Figures 6C and 6D). Because of the small number of the *H3*^{MUT} tumors (n = 7 and 12 in the discovery and second datasets, respectively), further verification is warranted.

Although the factors driving IDH1/2 protein abundance in these pediatric HGG tumors remain largely unknown, one possible factor is revealed by the CNV-RNA/protein *cis*-regulation investigation, which identified IDH1 as a CNV-RNA/protein cascade protein (Figure S3D). As illustrated in Figure S6D, *IDH1* deletion, which was observed in about 20% of HGG tumors, significantly downregulated the protein abundance of IDH1.

To nominate potential drug targets for pediatric HGG based on new insights from proteogenomics data, we performed a drug connectivity analysis to identify drug candidates whose effect on the transcriptome and proteome is diametrically opposed to the characteristics identified as central to HGG biology. Because of the lack of adjacent normal tissue of individuals with HGG, we chose to derive RNA/proteomics signatures of HGG aggressiveness by comparing HGG and LGG tumors. We then leveraged the LINCS L1000 transcriptomics and P100 phosphoproteomics perturbation-response databases to search for candidate drugs inducing effects that oppose the corresponding input (Litichevskiy et al., 2018; Subramanian et al., 2017; STAR Methods; Table S6). CDK inhibitors were predicted to reverse the aggressiveness of HGG based on RNA and phosphoproteomics data (Figures 6G, S6F, and S6G; Table S6). The kinase activity of CDK1 and CDK2 was upregulated in HGG compared with LGG (Figure 6H). MEK, proteasome, and HDAC inhibitors were found to be significant based on RNA data alone (Figures 6G, S6F, and S6G). Although MEK substrates were not observed in the phosphoproteomics data of HGG samples, *MAPK1* kinase activity, downstream of MEK, was found to be upregulated in HGG tumors (Figure 6H). Upregulation of substrates of CDK and MAPK1 proteins in HGG tumors supports that CDK inhibitors and MEKis might be effective for HGG tumors.

Comparison between Initial and Recurrent Tumors

Earlier work has reported distinct patterns between initial and recurrent tumors of the same individual (Morrissy et al., 2016). Based on proteomics and genomics profiles of 18 pairs of surgical samples from two distinct disease occurrences of the same individuals in our cohort, we tried to address the question of whether the recurrent tumors should be considered independent tumors during treatment evaluation.

In all 18 pairs, the recurrent/progressive (RP) disease was of the same histological diagnosis as the initial (IN) tumor. RP tumors carried 0%–52% (mean, 18%) of the IN tumor mutations (Figure 7A), which was lower than that of adult GBM (Cancer Genome Atlas Research Network, 2008) and LGG (Brat et al., 2015; Figure S7A). Remarkably, all three MB progression samples had a *TP53* mutation that was absent in their paired primary tumors (Figure 7A), consistent with the observation by Morrissy et al. (2016). In addition, there was an increase in chromosome arm aberrations in the RP samples, with the number of breakpoints increasing, on average, from 32 to 53 (Figure 7A). In contrast, in adult GBM and LGG tumors, CNV events from primary and recurrent tumors were similar (Figure S7A). Proteomics profiles also revealed differences between RP and IN tumors. In fact, most primary and recurrent samples were classified into different proteomics clusters (Figure 7A). As an example, one *BRAF*^{WT} LGG case (pair 173.2154), with BRAF being wild-type in its IN and RP tumors, had its IN tumor allocated to the *LGG BRAF*^{WT}*-rich* proteomic cluster and the RP tumor allocated to the *Cranio/LGG BRAF*^{V600E} cluster. Consistent with the characteristics of these two proteomics clusters, we observed upregulation of RNA transcription and splicing/EMT and coagulation and downregulation of the gap junction in RP tumors compared with their IN counterparts (Figures 1D and 7A; Table S7). Consistent with the allocation of RP tumors to the *Cranio/LGG BRAF*^{V600E} cluster, higher kinase activity for ERK1/ERK2 was observed in RP tumors compared with IN tumors (Figure 7C), suggesting that MEKi therapy might be more beneficial for RP tumors. Another LGG case (pair 350.944) had the IN tumor allocated to *Cranio/LGG BRAF*^{V600E}, whereas the RP sample was allocated to the *LGG BRAF*^{WT}*-rich* cluster, which resulted in an opposite trend in pathway activities, followed by a reversed trend in the activity of ERK1/ERK2 (Figures 7A and 7C). These changes in pathway activation highlight the need for *de novo* characterization of recurrent cases, which might affect treatment decisions.

We also investigated the correlation between IN and RP proteome profile pairs (Figures 7B and S7B). The fact that a good number of primary-recurrent pairs were not highly correlated with each other at the proteomics level supports the idea that recurrence (or progression) of a tumor could have different tumorigenesis mechanisms. Based on these observations, an approach that assesses the molecular properties of recurrent events independent of the IN tumor seems to be warranted.

DISCUSSION

This study represents the very first attempt to perform a large-scale proteogenomics-integrative analysis for multiple distinct pediatric brain tumor diagnoses to discover new effective targeted therapies. High-quality genomics, transcriptomics, proteomics, and phosphoproteomics data were generated as a public

(B) Distribution of Spearman's correlation between the proteomic abundance of any pair of tumors within a particular histology. Correlations between the 18 paired IN-RP samples were further labeled in the violin plots.

(C) Distribution of kinase activity scores of MAPK1/3 among all LGG samples, LGG samples allocated to C4, and LGG samples allocated to C8. IN and RP samples of individuals LGG.350.944 and LGG.173.2154 are highlighted.

See also Figure S7.

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

resource from a retrospective cohort of 218 frozen tissue samples collected by a single institution.

Although histological diagnosis remains the cornerstone of classifying tumors into therapeutic categories, it is now well recognized that molecular subgroups within histologically similar tumors can be identified on the basis of transcriptomics, genomics, and methylomics. Our study is based on the recognition that proteomics/phosphoproteomics need to be integrated with other omics to gain an improved system biology view of molecular subgroups. In addition, we advocate the importance of characterizing biological themes that cross histological boundaries and unite individual tumors of disparate histologies and cells of origin because such insights can lead to new extensions of treatments shown to be effective in one type of tumor to other, histologically disparate tumors sharing the same proteomics features. For example, our proteomics/phosphoproteomics data clustering analyses revealed two distinct subgroups of pediatric CP, with one subgroup showing strikingly similar proteomics/phosphoproteomics characteristics as pediatric LGG *BRAF*^{V600E} tumors. This observation suggests potential use of MEK/MAPK inhibitors in a subset of pediatric CP, which currently has no robust chemotherapy options.

The existence of two subgroups of pediatric CP, however, is not evident from RNA data. Similarly, we observed profound discordance between RNA and protein abundance in other histologies, such as EP and LGG. The low/moderate RNA-protein correlation observed in this project is consistent with other large-scale proteogenomics projects (Clark et al., 2019; Dou et al., 2020; Gillette et al., 2020). Multiple factors, such as protein turnover and selective translation, contribute to the low correlation between RNA and protein abundance (Clark et al., 2019; Dou et al., 2020). Interestingly, more aggressive tumors tend to show increased protein-RNA correlation, a phenomenon observed across multiple cancer proteogenomics studies (Clark et al., 2019; Dou et al., 2020). One possible explanation is that aggressive tumors often have high proliferation, and the boosted translation activities in highly proliferative (tumor) cells result in more correlated RNA and protein signals. Thus, studying the proteome reveals insights not evident from RNA-based analysis alone.

There is also significant value in integration of large-scale proteomics and genomics data to identify the ramifications of genomic events on biological function. A good illustration of this contribution is the ability to discern, at the protein level, the *cis* effects of copy number alterations by tracking the cascade of abundance from gene dose to transcript level to protein/phosphosite abundance. In this way, the relevant genes in a chromosomal region with altered copy number can be identified for validation of their biologic contribution (e.g., *RAGBAP1L* in EP and *FDPS* in HGG with *1q* gain).

It must be noted that a clear view of the proteomics contribution needs to acknowledge a presumed equivalence between abundance and activity. However, investigation of kinase activity based on phosphoproteomics showed that protein abundance can be reduced in active signaling pathways (e.g., ERK1/3 in CP), which could happen because of feedback loops in many complicated regulatory processes, suggesting the important role of phosphoproteomics data in characterizing pathway activities.

Many molecularly targeted agents have now been sufficiently characterized in terms of their safety and mechanisms of action to allow combinations of these agents to enter clinical trials. Selective targeting of multiple kinase nodes in these networks is a common strategy to construct more effective treatments while leaving fewer redundant escape routes for the tumor cell. Concurrently targeting the MEK and mTOR pathways in RAF-activated tumors is a strategy that has emerged from investigations of resistance pathways resulting from RAF/MAPK inhibition. Our characterization of proteome and phosphoproteome changes because of *BRAF* aberrations lends further rationale to this approach in LGG, where resistance per se is not an issue but durability of response after treatment cessation is.

The retrospective study design enabled us to access follow-up clinical data, including outcomes. This strategy also allowed us to study rare diagnoses and compare primary and recurrent tumors from the same individuals. One interesting finding was that the protein abundance of the wild-type IDH1/2 is prognostic in HGG tumors without H3K27M mutation. In addition, by comparing primary and recurrent specimens, we were able to discern significant differences between these paired specimens. The shifts in underlying biology accompanying tumor recurrence necessitate independent assessment and therapy decisions for those recurrent tumors.

This project represents a significant advance in biological interrogation of pediatric brain tumors at multiple levels of biological control and across traditional histological boundaries. Although the limited sample sizes of some histologies pose a significant limitation for certain investigations, as the first proteogenomic characterization of such histologies, the data and analytical results from this project serve as a valuable resource. Importantly, it is the result of a necessarily expansive partnership between children's hospitals, patients and families, philanthropic and federal funding, and physician scientists and computational biologists. Such endeavors demonstrate the potential of large-scale proteogenomics science and the power of inclusive collaboration to tackle a pervasive threat to our children: pediatric brain tumors.

Limitations and Future Directions

Although this project represents the most comprehensive multi-omics analysis of pediatric brain tumors ever undertaken, there are nevertheless a number of limitations that are the result of the rarity of the tumor types studied and the nature of the samples available. (1) There are additional layers of cellular regulation not included in this initiative, such as methylation profiling, histone mark profiling, ribosome sequencing, metabolomics, and acetyl-proteomics. (2) This study was structured within tissue access limitations to provide the ability to discern common biology across major histological types of pediatric brain cancer. Therefore, we sacrificed the ability to perform in-depth proteomic analyses within the tumor types that were represented by smaller sample sizes. A future study gathering larger cohorts of the less common brain tumor types would be instructive in identifying the biology that is unique to those tumors. (3) This study leveraged retrospective tissue collection, which allowed analysis of survival outcome and IN/progressive tumor pairs and made it

feasible to study rare tumors. The cost of this approach, however, was that the samples were not fully exploited for phospho-proteomics, and our phosphosite data, although significant and high quality, considering the amount of tissue available, would have been deeper if we had samples that had been acquired prospectively with a phospho-proteomics-specific collection protocol.

Acknowledging these limitations, we can also see some tantalizing future possibilities made more clear by the findings of this project. (1) Our study demonstrates the ability of proteomics, phospho-proteomics, and kinase activity scores to elucidate active signaling processes within tumors. Applying these capabilities to tissue samples in a clinical trial context could yield valuable information regarding the biology of individual tumors that respond to a given therapy. (2) As these determinant proteomic signatures are identified, MRM signatures can be developed to identify individuals in real time whose tumors display particular biologic features and, thus, may respond to a treatment. (3) Histologically similar tumors are frequently treated differently in pediatric and adult settings. Although they often differ in their genomic features, this study has shown that they do not always drive biology. A unique opportunity building on this work will be to use proteomics platforms to interrogate tumors whose incidence spans a large age range to answer questions regarding how biology changes across the spectrum and whether treatments can be realigned for maximum benefit. In summary, this large multi-omics study of pediatric brain tumors represents an entrée for integration of proteomics into data science modeling of pediatric cancer; therefore, it sets the stage for more applied research to come.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Patient selection for the discovery cohort
  - Sample collection for the discovery cohort
  - Sample collection for the HGG validation study
- METHOD DETAILS
  - Nucleic acid extractions, WGS and RNaseq
  - Somatic Mutation and CNV calling
  - RNaseq data preprocessing
  - Proteomic experiments for Discovery Cohort
  - Protein Identification and Quantification
  - Preprocessing of TMT proteomic data
  - QC check for proteogenomic profiles
  - Proteomics experiment of the validation cohort
  - Targeted Mass Spectrometry Methods
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Kinase Activity Score Calculation
  - Consensus clustering analysis
  - Survival analysis for proteomic clusters
  - Stemness Score
  - Proliferative Index
  - Investigation of two subtypes of Cranio
  - Proteomic Cluster Signature
  - Pathway analysis for proteomic clusters
  - Immune subtype identification
  - single cell RNaseq deconvolution Analysis
  - Tumor Purity, Stromal and Immune Scores
  - Differentially Expressed Genes and Pathway
  - Microglia and Macrophage Polarization in LGG
  - Immune association with BRAF Status in LGG
  - iProFun Based Cis Association Analysis
  - Cis-regulation of Somatic Mutations
  - Trans association analysis
  - Kinase Activity across different histologies
  - Kinase-Substrates Association
  - Validation of kinase-phospho associations
  - BRAF mutation association analysis for LGG
  - Pathways associated with BRAF status in LGG
  - Phosphoproteomic Co-expression Network in LGG
  - Pathway analysis of network modules
  - Survival analysis of HGG
  - Drug Connectivity Analysis for HGG
  - Comparison of initial and progressed tumors
  - Germline variants in TP53
- ADDITIONAL RESOURCES
  - Heatmap Web Server

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cell.2020.10.044.

CONSORTIA

The members of the National Cancer Institute Clinical Proteomic Tumor Analysis Consortium are Alexey I. Nesvizhskii, Alicia Francis, Amanda G. Paulovich, Ana I. Robles, Anna Calinawan, Antonio Iavarone, Antonio Colaprico, Arul M. Chinnaiyan, Azra Krek, Bing Zhang, Boris Reva, Christopher R. Kinsinger, D.R. Mani, David Fenyö, Dmitry Rykunov, Emily Boja, Emily A. Kawaler, Eric E. Schadt, Felipe da Veiga Leprevost, Francesca Petralia, Gad Getz, Gonzalo Lopez, Guo Ci Teo, Henry Rodriguez, Hui-Yin Chang, Jacob J. Kennedy, Jamie Moon, Jeffrey R. Whiteaker, Jiayi Ji, John Spytz, Joshua Wang, Jun Zhu, Karen A. Ketchum, Karin D. Rodland, Karl K. Weitz, Karsten Krug, Kelly V. Ruggles, Lei Zhao, Li Ding, Liang-Bo Wang, Lili Blumenberg, Lizabeth Katsnelson, Maciej Wiznerowicz, MacIntosh Cornwell, Marcin J. Domagalski, Marcin P. Cieślik, Marina A. Gritsenko, Matilda Broberg, Matthew J. Ellis, Matthew A. Wyczalkowski, Matthew E. Monroe, Mehdi Mesri, Nathan J. Edward, Nicole Tignor, Pei Wang, Peter B. McGarvey, Richard G. Ivey, Richard D. Smith, Ronald J. Moore, Rosalie K. Chu, Rui Zhao, Sanjukta Guha Thakurta, Saravana Mohan Dhanasekaran, Selim Kalayci, Seungyeul Yoo, Shrabanti Chowdhury, Shuang Cai, Stephan Schürer, Steven P. Gygi, Tao Liu, Tara Hiltke, Tatiana Omelchenko, Toan Le, Travis D. Lorentzen, Uliana J. Voytovich, Vasileios Stathias, Vladislav A. Petyuk, Weiping Ma, Wenke Liu, William E. Bocik, Xi Steven Chen, Xiaoyu Song, Ying Wang, and Zeynep Gümüş.

The members of the Children's Brain Tumor Network (CBTN) are Adam C. Resnick, Nithin D. Adappa, Elizabeth Appert, Valerie Baubet, Oren Becher, Miguel A. Brown, Anna Maria Buccoliero, Jason E. Cain, Chiara Caporalini, Meghan Connors, Brian M. Ennis, Bailey Farrow, Ron Firestein, Jessica B. Foster, Krutika S. Gaonkar, Gerald Grant, Jeffrey P. Greenfield, Yiran Guo, Nalin Gupta, Derek Hanson, Allison P. Heath, Eric M. Jackson, Cassie N. Kline,

Mateusz Koptyra, David E. Kram, Sarah Leary, Carina A. Leonard, Chunde Li, Marilyn M. Li, Jena V. Lilly, Robert M. Lober, Jennifer L. Mason, Allison M. Morgan, Sabine Mueller, Javad Nazarian, Judy E. Palma, James Palmer, Sonia Partap, Tatiana Patton, Joanna J. Phillips, Ian F. Pollack, Eric H. Raabe, Pichai Raman, Samuel Rivero-Hinojosa, Jo Lynne Rokita, Brian R. Rood, Mariarita Santi, Mahdi Sarmady, Mirko Scagnet, Gabrielle S. Stone, Phillip B. Storm, Lamiya Tauhid, Olena M. Vaske, Angela N. Viaene, Nicholas A. Vitanza, Angela J. Waanders, Samuel G. Winebrake, Nathan Young, Bo Zhang, and Yuankun Zhu.

ACKNOWLEDGMENTS

This work was supported by the National Cancer Institute (NCI) Clinical Proteomic Tumor Analysis Consortium (CPTAC) grants U24CA210993, U01 CA214114, U24CA210967, U24CA210954, U24CA210972, U24CA210955, U24CA210979, and R50 CA211499. We would like to thank the children and their families for donating tumor samples for this study, the Children's Brain Tumor Network, and Dr. David Stokes and the entire Biorepository Resource Center (BioRC) at Children's Hospital of Philadelphia. This work was supported in part through the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai. The MS-based proteomic analysis of the Project Hope brain tumor samples was performed at the Environmental Molecular Sciences Laboratory, a U.S. Department of Energy National Scientific User Facility located at the Pacific Northwest National Laboratory in Richland, WA, operated by the Battelle Memorial Institute for the DOE under contract DE-AC05-76RL01830.

AUTHOR CONTRIBUTIONS

Study Conception & Design, H.R., A.C.R., A.G.P., S.P.G., P.B.S., B.R.R., and P.W.; Experiments and Data Collection, M.K., Y.Z., A.P.H., J.L.R., J.L.M., J.V.L., B.F., N.Y., A.C.R., S.P.G., S.G.T., J.S., A.G.P., U.J.V., L.Z., J.J.K., R.G.I., J.R.W., M.A.G., K.K.W., J.L.M., V.A.P., R.K.C., R.J.M., M.E.M., R.Z., R.D.S., K.D.R., and T.L.; Data Analysis, M.A.B., Y.G., K.S.G., P.R., B.M.E., B.Z., Y.Z., F.d.V.L., H.-Y.C., A.I.N., F.P., A.I., N.T., S.C., W.M., J.J., X.S., B.R., D.R., A.K., P.W., G.L., S.Y., J.Z, V.S., S.S., A. Colaprico., X.S.C., B.R., A.C.R., D.F., M.B., J.M.W., Y.W., L.-B.W., M.W., and L.K.; Sample Resources, A.C.R., J.V.L., J.L.M., P.B.S., V.B., and L.T.; Writing, F.P, N.T., M.K., S.C., W.M., M.A.B., K.S.G., J.L.R., D.R., A.K., X.S., J.J., S.Y., A.C.R., B.R.R., and P.W.; Visualization, S.K., G.L., Z.H.G., A. Calinawan, F.P., D.R., A.K., N.T., W.M., S.C., X.S., and J.J.; Supervision, P.W., B.R.R., A.C.R., A.G.P., S.P.G., J.L.R., Y.Z., M.K., A.P.H., P.B.S., and D.F.; Project Administration, P.W., H.R., E.B., T.H., M.M., F.P., N.T., B.R.R., and A.C.R.; Funding Acquisition, P.W., B.R.R., P.B.S., A.C.R., A.G.P., S.P.G., T.L., R.D.S., D.F., A.I.N., L.D., D.R.M., E.E.S., B.Z., and H.R. All authors contributed to data interpretation, manuscript editing, and revision.

DECLARATION OF INTERESTS

E.E.S. serves as chief executive officer for Sema4 and has an equity interest in this company.

Received: December 18, 2019
Revised: June 19, 2020
Accepted: October 26, 2020
Published: November 25, 2020

REFERENCES

Abate, M., Laezza, C., Pisanti, S., Torelli, G., Seneca, V., Catapano, G., Montella, F., Ranieri, R., Notarnicola, M., Gazzerro, P., et al. (2017). Deregulated expression and activity of Farnesyl Diphosphate Synthase (FDPS) in Glioblastoma. Sci. Rep. *7*, 14123.

Alyass, A., Turcotte, M., and Meyre, D. (2015). From big data analysis to personalized medicine for all: challenges and opportunities. BMC medical genomics *8*, 1–25.

Apps, J.R., Carreno, G., Gonzalez-Meljem, J.M., Haston, S., Guiho, R., Cooper, J.E., Manshaei, S., Jani, N., Hölsken, A., Pettorini, B., et al. (2018). Tumour compartment transcriptomics demonstrates the activation of inflammatory and odontogenic programmes in human adamantinomatous craniopharyngioma and identifies the MAPK/ERK pathway as a novel therapeutic target. Acta Neuropathol. *135*, 757–777.

Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. *18*, 220.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.; The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29.

Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al.; MC3 Working Group; Cancer Genome Atlas Research Network (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell *173*, 371–385.e18.

Beausoleil, S.A., Villén, J., Gerber, S.A., Rush, J., and Gygi, S.P. (2006). A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nat. Biotechnol. *24*, 1285–1292.

Bednarek, E., and Caroni, P. (2011). β-Adducin is required for stable assembly of new synapses and improved memory upon environmental enrichment. Neuron *69*, 1132–1146.

Behrens, J., Kameritsch, P., Wallner, S., Pohl, U., and Pogoda, K. (2010). The carboxyl tail of Cx43 augments p38 mediated cell migration in a gap junction-independent manner. Eur. J. Cell Biol. *89*, 828–838.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. B Methodol. *57*, 289–300.

Bergaggio, E., and Piva, R. (2019). Wild-Type IDH Enzymes as Actionable Targets for Cancer Therapy. Cancers (Basel) *11*, 563.

Box, J.K., Paquet, N., Adams, M.N., Boucher, D., Bolderson, E., O'Byrne, K.J., and Richard, D.J. (2016). Nucleophosmin: from structure and function to disease development. BMC Mol. Biol. *17*, 19.

Brastianos, P.K., Taylor-Weiner, A., Manley, P.E., Jones, R.T., Dias-Santagata, D., Thorner, A.R., Lawrence, M.S., Rodriguez, F.J., Bernardo, L.A., Schubert, L., et al. (2014). Exome sequencing identifies BRAF mutations in papillary craniopharyngiomas. Nat. Genet. *46*, 161–165.

Brat, D.J., Verhaak, R.G., Aldape, K.D., Yung, W.K., Salama, S.R., Cooper, L.A., Rheinbay, E., Miller, C.R., Vitucci, M., Morozova, O., et al.; Cancer Genome Atlas Research Network (2015). Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. N. Engl. J. Med. *372*, 2481–2498.

Butovsky, O., Jedrychowski, M.P., Moore, C.S., Cialic, R., Lanser, A.J., Gabriely, G., Koeglsperger, T., Dake, B., Wu, P.M., Doykan, C.E., et al. (2014). Identification of a unique TGF-β-dependent molecular and functional signature in microglia. Nat. Neurosci. *17*, 131–143.

Cai, Z., Wang, Y., Yu, W., Xiao, J., Li, Y., Liu, L., Zhu, C., Tan, K., Deng, Y., Yuan, W., et al. (2006). hnulp1, a basic helix-loop-helix protein with a novel transcriptional repressive domain, inhibits transcriptional activity of serum response factor. Biochem. Biophys. Res. Commun. *343*, 973–981.

Cai, Y., Guo, T., Wang, Y., and Du, J. (2018). Glutamate Metabolism Regulates Immune Escape of Glioma. Madridge Journal of Immunology *2*, 53–57.

Calvert, A.E., Chalastanis, A., Wu, Y., Hurley, L.A., Kouri, F.M., Bi, Y., Kachman, M., May, J.L., Bartom, E., Hua, Y., et al. (2017). Cancer-Associated IDH1 Promotes Growth and Resistance to Targeted Therapies in the Absence of Mutation. Cell Rep. *19*, 1858–1873.

Campanini, M.L., Colli, L.M., Paixao, B.M., Cabral, T.P., Amaral, F.C., Machado, H.R., Neder, L.S., Saggioro, F., Moreira, A.C., Antonini, S.R., and de Castro, M. (2010). CTNNB1 gene mutations, pituitary transcription factors, and MicroRNA expression involvement in the pathogenesis of adamantinomatous craniopharyngiomas. Horm. Cancer *1*, 187–196.

Cancer Genome Atlas Research Network (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature *455*, 1061–1068.

Capper, D., Jones, D.T.W., Sill, M., Hovestadt, V., Schrimpf, D., Sturm, D., Koelsche, C., Sahm, F., Chavez, L., Reuss, D.E., et al. (2018). DNA methylation-based classification of central nervous system tumours. Nature *555*, 469–474.

Chalmers, Z.R., Connelly, C.F., Fabrizio, D., Gay, L., Ali, S.M., Ennis, R., Schrock, A., Campbell, B., Shlien, A., Chmielecki, J., et al. (2017). Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Med. *9*, 34.

Chang, C.V., Araujo, R.V., Cirqueira, C.S., Cani, C.M., Matushita, H., Cescato, V.A., Fragoso, M.C., Bronstein, M.D., Zerbini, M.C., Mendonca, B.B., and Carvalho, L.R. (2017). Differential Expression of Stem Cell Markers in Human Adamantinomatous Craniopharyngioma and Pituitary Adenoma. Neuroendocrinology *104*, 183–193.

Chen, J.H., Huang, S.M., Chen, C.C., Tsai, C.F., Yeh, W.L., Chou, S.J., Hsieh, W.T., and Lu, D.Y. (2011). Ghrelin induces cell migration through GHS-R, CaMKII, AMPK, and NF-κB signaling pathway in glioma cells. J. Cell. Biochem. *112*, 2931–2941.

Chisci, E., De Giorgi, M., Zanfrini, E., Testasecca, A., Brambilla, E., Cinti, A., Farina, L., Kutryb-Zajac, B., Bugarin, C., Villa, C., et al. (2017). Simultaneous overexpression of human E5NT and ENTPD1 protects porcine endothelial cells against H_2O_2-induced oxidative stress and cytotoxicity in vitro. Free Radic. Biol. Med. *108*, 320–333.

Clark, D.J., Dhanasekaran, S.M., Petralia, F., Pan, J., Song, X., Hu, Y., da Veiga Leprevost, F., Reva, B., Lih, T.M., Chang, H.Y., et al.; Clinical Proteomic Tumor Analysis Consortium (2019). Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. Cell *179*, 964–983.e31.

Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S., Malta, T.M., Pagnotta, S.M., Castiglioni, I., et al. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. *44*, e71.

Colaprico, A., Olsen, C., Bailey, M.H., Odom, G.J., Terkelsen, T., Silva, T.C., Olsen, A.V., Cantini, L., Zinovyev, A., Barillot, E., et al. (2020). Interpreting pathways to discover cancer driver genes with Moonlight. Nat. Commun. *11*, 69.

Cole, A.R. (2012). GSK3 as a Sensor Determining Cell Fate in the Brain. Front. Mol. Neurosci. *5*, 4.

Cooper, C.D., and Lampe, P.D. (2002). Casein kinase 1 regulates connexin-43 gap junction assembly. J. Biol. Chem. *277*, 44962–44968.

Coy, S., Rashid, R., Lin, J.R., Du, Z., Donson, A.M., Hankinson, T.C., Foreman, N.K., Manley, P.E., Kieran, M.W., Reardon, D.A., et al. (2018). Multiplexed immunofluorescence reveals potential PD-1/PD-L1 pathway vulnerabilities in craniopharyngioma. Neuro-oncol. *20*, 1101–1112.

Crotti, A., and Ransohoff, R.M. (2016). Microglial Physiology and Pathophysiology: Insights from Genome-wide Transcriptional Profiling. Immunity *44*, 505–515.

Csárdi, G., and Nepusz, T. (2006). The igraph software package for complex network research. InterJournal Complex Systems, 1695.

Cuddapah, V.A., and Sontheimer, H. (2010). Molecular interaction and functional regulation of ClC-3 by Ca2+/calmodulin-dependent protein kinase II (CaMKII) in human malignant glioma. J. Biol. Chem. *285*, 11188–11196.

da Veiga Leprevost, F., Haynes, S.E., Avtonomov, D.M., Chang, H.Y., Shanmugam, A.K., Mellacheruvu, D., Kong, A.T., and Nesvizhskii, A.I. (2020). Philosopher: a versatile toolkit for shotgun proteomics data analysis. Nat. Methods *17*, 869–870.

Daily, K., Ho Sui, S.J., Schriml, L.M., Dexheimer, P.J., Salomonis, N., Schroll, R., Bush, S., Keddache, M., Mayhew, C., Lotia, S., et al. (2017). Molecular, phenotypic, and sample-associated data to describe pluripotent stem cell lines and derivatives. Sci. Data *4*, 170030.

Darmanis, S., Sloan, S.A., Croote, D., Mignardi, M., Chernikova, S., Samghababi, P., Zhang, Y., Neff, N., Kowarsky, M., Caneda, C., et al. (2017). Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. Cell Rep. *21*, 1399–1410.

Dello Russo, C., Lisi, L., Tentori, L., Navarra, P., Graziani, G., and Combs, C.K. (2017). Exploiting Microglial Functions for the Treatment of Glioblastoma. Curr. Cancer Drug Targets *17*, 267–281.

Djomehri, S.I., Gonzalez, M.E., da Veiga Leprevost, F., Tekula, S.R., Chang, H.Y., White, M.J., Cimino-Mathews, A., Burman, B., Basrur, V., Argani, P., et al. (2020). Quantitative proteomic landscape of metaplastic breast carcinoma pathological subtypes and their relationship to triple-negative tumors. Nat. Commun. *11*, 1723.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dou, Y., Kawaler, E.A., Cui Zhou, D., Gritsenko, M.A., Huang, C., Blumenberg, L., Karpova, A., Petyuk, V.A., Savage, S.R., Satpathy, S., et al.; Clinical Proteomic Tumor Analysis Consortium (2020). Proteogenomic Characterization of Endometrial Carcinoma. Cell *180*, 729–748.e26.

Easley-Neal, C., Fierro, J., Jr., Buchanan, J., and Washbourne, P. (2013). Late recruitment of synapsin to nascent synapses is regulated by Cdk5. Cell Rep. *3*, 1199–1212.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The Reactome Pathway Knowledgebase. Nucleic Acids Res. *46* (D1), D649–D655.

Fangusaro, J., Onar-Thomas, A., Young Poussaint, T., Wu, S., Ligon, A.H., Lindeman, N., Banerjee, A., Packer, R.J., Kilburn, L.B., Goldman, S., et al. (2019). Selumetinib in paediatric patients with BRAF-aberrant or neurofibromatosis type 1-associated recurrent, refractory, or progressive low-grade glioma: a multicentre, phase 2 trial. Lancet Oncol. *20*, 1011–1022.

Farghaian, H., Turnley, A.M., Sutherland, C., and Cole, A.R. (2011). Bioinformatic prediction and confirmation of beta-adducin as a novel substrate of glycogen synthase kinase 3. J. Biol. Chem. *286*, 25274–25283.

Fasolini, M., Wu, X., Flocco, M., Trosset, J.Y., Oppermann, U., and Knapp, S. (2003). Hot spots in Tcf4 for the interaction with beta-catenin. J. Biol. Chem. *278*, 21092–21098.

Fei, L., and Xu, H. (2018). Role of MCM2-7 protein phosphorylation in human cancer cells. Cell Biosci. *8*, 43.

Fruchterman, T.M.J., and Reingold, E.M. (1991). Graph drawing by force-directed placement. Softw. Pract. Exper. *21*, 1129–1164.

Fumagalli, M., Lombardi, M., Gressens, P., and Verderio, C. (2018). How to reprogram microglia toward beneficial functions. Glia *66*, 2531–2549.

Gao, C., Wang, Y., Broaddus, R., Sun, L., Xue, F., and Zhang, W. (2017). Exon 3 mutations of *CTNNB1* drive tumorigenesis: a review. Oncotarget *9*, 5492–5508.

Gibbons, B.C., Chambers, M.C., Monroe, M.E., Tabb, D.L., and Payne, S.H. (2015). Correcting systematic bias and instrument measurement drift with mzRefinery. Bioinformatics *31*, 3838–3840.

Gillette, M.A., Satpathy, S., Cao, S., Dhanasekaran, S.M., Vasaikar, S.V., Krug, K., Petralia, F., Li, Y., Liang, W.W., Reva, B., et al.; Clinical Proteomic Tumor Analysis Consortium (2020). Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma. Cell *182*, 200–225.e35.

Gröbner, S.N., Worst, B.C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V.A., Johann, P.D., Balasubramanian, G.P., Segura-Wang, M., Brabetz, S., et al.; ICGC PedBrain-Seq Project; ICGC MMML-Seq Project (2018). The landscape of genomic alterations across childhood cancers. Nature *555*, 321–327.

Haage, V., Semtner, M., Vidal, R.O., Hernandez, D.P., Pong, W.W., Chen, Z., Hambardzumyan, D., Magrini, V., Ly, A., Walker, J., et al. (2019). Comprehensive gene expression meta-analysis identifies signature genes that distinguish microglia from peripheral monocytes/macrophages in health and glioma. Acta Neuropathol. Commun. *7*, 20.

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics *14*, 7.

Hayashi, H., Matsuzaki, O., Muramatsu, S., Tsuchiya, Y., Harada, T., Suzuki, Y., Sugano, S., Matsuda, A., and Nishida, E. (2006). Centaurin-alpha1 is a phosphatidylinositol 3-kinase-dependent activator of ERK1/2 mitogen-activated protein kinases. J. Biol. Chem. *281*, 1332–1337.

Hong, X., Sin, W.C., Harris, A.L., and Naus, C.C. (2015). Gap junctions modulate glioma invasion by direct transfer of microRNA. Oncotarget *6*, 15566–15577.

Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., and Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. *43*, D512–D520.

Itoh, T., Satoh, M., Kanno, E., and Fukuda, M. (2006). Screening for target Rabs of TBC (Tre-2/Bub2/Cdc16) domain-containing proteins based on their Rab-binding activity. Genes Cells *11*, 1023–1037.

Jain, P., Silva, A., Han, H.J., Lang, S.S., Zhu, Y., Boucher, K., Smith, T.E., Vakil, A., Diviney, P., Choudhari, N., et al. (2017). Overcoming resistance to single-agent therapy for oncogenic *BRAF* gene fusions *via* combinatorial targeting of MAPK and PI3K/mTOR signaling pathways. Oncotarget *8*, 84697–84713.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127.

Kalayci, S., and Gumus, Z.H. (2018). Exploring Biological Networks in 3D, Stereoscopic 3D, and Immersive 3D with iCAVE. Curr. Protoc. Bioinformatics *61*, 8.27.21–28.27.26.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. *45* (D1), D353–D361.

Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. Nature *581*, 434–443.

Karmakar, S., Dey, P., Vaz, A.P., Bhaumik, S.R., Ponnusamy, M.P., and Batra, S.K. (2018). PD2/PAF1 at the Crossroads of the Cancer Network. Cancer Res. *78*, 313–319.

Karremann, M., Gielen, G.H., Hoffmann, M., Wiese, M., Colditz, N., Warmuth-Metz, M., Bison, B., Claviez, A., van Vuurden, D.G., von Bueren, A.O., et al. (2018). Diffuse high-grade gliomas with H3 K27M mutations carry a dismal prognosis independent of tumor location. Neuro-oncol. *20*, 123–131.

Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. *74*, 5383–5392.

Kilday, J.P., Mitra, B., Domerg, C., Ward, J., Andreiuolo, F., Osteso-Ibanez, T., Mauguen, A., Varlet, P., Le Deley, M.C., Lowe, J., et al. (2012). Copy number gain of 1q25 predicts poor progression-free survival for pediatric intracranial ependymomas and enables patient risk stratification: a prospective European clinical trial cohort analysis on behalf of the Children's Cancer Leukaemia Group (CCLG), Societe Francaise d'Oncologie Pediatrique (SFOP), and International Society for Pediatric Oncology (SIOP). Clin. Cancer Res. *18*, 2001–2011.

Kim, W., and Liau, L.M. (2012). IDH mutations in human glioma. Neurosurg. Clin. N. Am. *23*, 471–480.

Kim, S., and Pevzner, P.A. (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. Nat. Commun. *5*, 5277.

Kim, S., Gupta, N., and Pevzner, P.A. (2008). Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. J. Proteome Res. *7*, 3354–3363.

Kim, H.Y., Kim, D.K., Bae, S.H., Gwak, H., Jeon, J.H., Kim, J.K., Lee, B.I., You, H.J., Shin, D.H., Kim, Y.H., et al. (2018a). Farnesyl diphosphate synthase is important for the maintenance of glioblastoma stemness. Exp. Mol. Med. *50*, 1–12.

Kim, S., Scheffler, K., Halpern, A.L., Bekritsky, M.A., Noh, E., Källberg, M., Chen, X., Kim, Y., Beyter, D., Krusche, P., and Saunders, C.T. (2018b). Strelka2: fast and accurate calling of germline and somatic variants. Nat. Methods *15*, 591–594.

Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D., and Nesvizhskii, A.I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. Nat. Methods *14*, 513–520.

Kordaß, T., Osen, W., and Eichmüller, S.B. (2018). Controlling the Immune Suppressor: Transcription Factors and MicroRNAs Regulating CD73/NT5E. Front. Immunol. *9*, 813.

Korotkevich, G., Sukhov, V., and Sergushichev, A. (2019). Fast gene set enrichment analysis. bioRxiv. https://doi.org/10.1101/060012.

Koschmann, C., Zamler, D., MacKay, A., Robinson, D., Wu, Y.M., Doherty, R., Marini, B., Tran, D., Garton, H., Muraszko, K., et al. (2016). Characterizing and targeting PDGFRA alterations in pediatric high-grade glioma. Oncotarget *7*, 65696–65706.

Koscielny, G., An, P., Carvalho-Silva, D., Cham, J.A., Fumis, L., Gasparyan, R., Hasan, S., Karamanis, N., Maguire, M., Papa, E., et al. (2017). Open Targets: a platform for therapeutic target identification and validation. Nucleic Acids Res. *45* (D1), D985–D994.

Koseki, J., Colvin, H., Fukusumi, T., Nishida, N., Konno, M., Kawamoto, K., Tsunekuni, K., Matsui, H., Doki, Y., Mori, M., and Ishii, H. (2015). Mathematical analysis predicts imbalanced IDH1/2 expression associates with 2-HG-inactivating β-oxygenation pathway in colorectal cancer. Int. J. Oncol. *46*, 1181–1191.

Krasemann, S., Madore, C., Cialic, R., Baufeld, C., Calcagno, N., El Fatimy, R., Beckers, L., O'Loughlin, E., Xu, Y., Fanek, Z., et al. (2017). The TREM2-APOE Pathway Drives the Transcriptional Phenotype of Dysfunctional Microglia in Neurodegenerative Diseases. Immunity *47*, 566–581.e9.

Lakka, S.S., and Rao, J.S. (2008). Antiangiogenic therapy in brain tumors. Expert Rev. Neurother. *8*, 1457–1473.

Lee, S., Lee, S., Ouellette, S., Park, W.Y., Lee, E.A., and Park, P.J. (2017). NGSCheckMate: software for validating sample identity in next-generation sequencing studies within and across data types. Nucleic Acids Res. *45*, e103.

Leone, R.D., and Emens, L.A. (2018). Targeting adenosine for cancer immunotherapy. J. Immunother. Cancer *6*, 57.

Leone, R.D., Zhao, L., Englert, J.M., Sun, I.M., Oh, M.H., Sun, I.H., Arwood, M.L., Bettencourt, I.A., Patel, C.H., Wen, J., et al. (2019). Glutamine blockade induces divergent metabolic programs to overcome tumor immune evasion. Science *366*, 1013–1021.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, J., Davidson, D., Martins Souza, C., Zhong, M.C., Wu, N., Park, M., Muller, W.J., and Veillette, A. (2015). Loss of PTPN12 Stimulates Progression of ErbB2-Dependent Breast Cancer by Enhancing Cell Survival, Migration, and Epithelial-to-Mesenchymal Transition. Mol. Cell. Biol. *35*, 4069–4082.

Li, L., Xing, R., Cui, J., Li, W., and Lu, Y. (2018b). Investigation of frequent somatic mutations of MTND5 gene in gastric cancer cell lines and tissues. Mitochondrial DNA Part B *3*, 1002–1008.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739–1740.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425.

Liluashvili, V., Kalayci, S., Fluder, E., Wilson, M., Gabow, A., and Gümüs, Z.H. (2017). iCAVE: an open source tool for visualizing biomolecular networks in 3D, stereoscopic 3D and immersive 3D. Gigascience *6*, 1–13.

Litichevskiy, L., Peckner, R., Abelin, J.G., Asiedu, J.K., Creech, A.L., Davis, J.F., Davison, D., Dunning, C.M., Egertson, J.D., Egri, S., et al. (2018). A Library of Phosphoproteomic and Chromatin Signatures for Characterizing Cellular Responses to Drug Perturbations. Cell Syst. *6*, 424–443.e7.

Liu, H., Sun, Y., O'Brien, J.A., Franco-Barraza, J., Qi, X., Yuan, H., Jin, W., Zhang, J., Gu, C., Zhao, Z., et al. (2020). Necroptotic astrocytes contribute

to maintaining stemness of disseminated medulloblastoma through CCL2 secretion. Neuro-oncol. *22*, 625–638.

Lou, Y., Diao, L., Cuentas, E.R., Denning, W.L., Chen, L., Fan, Y.H., Byers, L.A., Wang, J., Papadimitrakopoulou, V.A., Behrens, C., et al. (2016). Epithelial-Mesenchymal Transition Is Associated with a Distinct Tumor Microenvironment Including Elevation of Inflammatory Signals and Multiple Immune Checkpoints in Lung Adenocarcinoma. Clin. Cancer Res. *22*, 3630–3642.

Lun, A.T., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Res. *5*, 2122.

Luke, J.J., Bao, R., Sweis, R.F., Spranger, S., and Gajewski, T.F. (2019). WNT/beta-catenin Pathway Activation Correlates with Immune Exclusion across Human Cancers. Clin. Cancer Res. *2*, 3074–3083.

Ma, W., Kim, S., Chowdhury, S., Li, Z., Yang, M., Yoo, S., Petralia, F., Jacobsen, J., Li, J.J., Ge, X., et al. (2020). DreamAI: algorithm for the imputation of proteomics data. bioRxiv. https://doi.org/10.1101/2020.07.21.214205.

Magupalli, V.G., Mochida, S., Yan, J., Jiang, X., Westenbroek, R.E., Nairn, A.C., Scheuer, T., and Catterall, W.A. (2013). Ca2+-independent activation of Ca2+/calmodulin-dependent protein kinase II bound to the C-terminal domain of CaV2.1 calcium channels. J. Biol. Chem. *288*, 4637–4648.

Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kamińska, B., Huelsken, J., Omberg, L., Gevaert, O., et al.; Cancer Genome Atlas Research Network (2018). Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. Cell *173*, 338–354.e15.

Meel, M.H., Schaper, S.A., Kaspers, G.J.L., and Hulleman, E. (2018). Signaling pathways and mesenchymal transition in pediatric high-grade glioma. Cell. Mol. Life Sci. *75*, 871–887.

Merlo, L.M., Pepper, J.W., Reid, B.J., and Maley, C.C. (2006). Cancer as an evolutionary and ecological process. Nat. Rev. Cancer *6*, 924–935.

Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2. 0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. *12*, https://doi.org/10.1186/gb-2011-12-4-r41.

Mertins, P., Yang, F., Liu, T., Mani, D.R., Petyuk, V.A., Gillette, M.A., Clauser, K.R., Qiao, J.W., Gritsenko, M.A., Moore, R.J., et al. (2014). Ischemia in tumors induces early and sustained phosphorylation changes in stress kinase pathways but does not affect global protein levels. Mol. Cell. Proteomics *13*, 1690–1704.

Mertins, P., Mani, D.R., Ruggles, K.V., Gillette, M.A., Clauser, K.R., Wang, P., Wang, X., Qiao, J.W., Cao, S., Petralia, F., et al.; NCI CPTAC (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. Nature *534*, 55–62.

Moniaux, N., Nemos, C., Schmied, B.M., Chauhan, S.C., Deb, S., Morikane, K., Choudhury, A., Vanlith, M., Sutherlin, M., Sikela, J.M., et al. (2006). The human homologue of the RNA polymerase II-associated factor 1 (hPaf1), localized on the 19q13 amplicon, is associated with tumorigenesis. Oncogene *25*, 3247–3257.

Monroe, M.E., Shaw, J.L., Daly, D.S., Adkins, J.N., and Smith, R.D. (2008). MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features. Comput. Biol. Chem. *32*, 215–217.

Monti, S., Tamayo, P., Mesirov, J.P., and Golub, T.R. (2003). Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. Machine Learning *52*, 91–118.

Morrissy, A.S., Garzia, L., Shih, D.J., Zuyderduyn, S., Huang, X., Skowron, P., Remke, M., Cavalli, F.M., Ramaswamy, V., Lindsay, P.E., et al. (2016). Divergent clonal selection dominates medulloblastoma at recurrence. Nature *529*, 351–357.

Mounir, M., Lucchetta, M., Silva, T.C., Olsen, C., Bontempi, G., Chen, X., Noushmehr, H., Colaprico, A., and Papaleo, E. (2019). New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. PLoS Comput. Biol. *15*, e1006701.

Navarrete-Perea, J., Yu, Q., Gygi, S.P., and Paulo, J.A. (2018). Streamlined Tandem Mass Tag (SL-TMT) Protocol: An Efficient Strategy for Quantitative (Phospho)proteome Profiling Using Tandem Mass Tag-Synchronous Precursor Selection-MS3. J. Proteome Res. *17*, 2226–2236.

Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. *75*, 4646–4658.

Newman, M.E., and Girvan, M. (2004). Finding and evaluating community structure in networks. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *69*, 026113.

Ng, E.L., and Tang, B.L. (2008). Rab GTPases and their roles in brain neurons and glia. Brain Res. Brain Res. Rev. *58*, 236–246.

Northcott, P.A., Buchhalter, I., Morrissy, A.S., Hovestadt, V., Weischenfeldt, J., Ehrenberger, T., Gröbner, S., Segura-Wang, M., Zichner, T., Rudneva, V.A., et al. (2017). The whole-genome landscape of medulloblastoma subtypes. Nature *547*, 311–317.

Ostrom, Q.T., Gittleman, H., Truitt, G., Boscia, A., Kruchko, C., and Barnholtz-Sloan, J.S. (2018). CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2011-2015. Neuro-oncol. *20* (suppl_4), iv1–iv86.

Parsons, D.W., Li, M., Zhang, X., Jones, S., Leary, R.J., Lin, J.C., Boca, S.M., Carter, H., Samayoa, J., Bettegowda, C., et al. (2011). The genetic landscape of the childhood cancer medulloblastoma. Science *331*, 435–439.

Pereira, M.S.L., Klamt, F., Thomé, C.C., Worm, P.V., and de Oliveira, D.L. (2017). Metabotropic glutamate receptors as a new therapeutic target for malignant gliomas. Oncotarget *8*, 22279–22298.

Perrot, I., Michaud, H.A., Giraudon-Paoli, M., Augier, S., Docquier, A., Gros, L., Courtois, R., Déjou, C., Jecko, D., Becquart, O., et al. (2019). Blocking Antibodies Targeting the CD39/CD73 Immunosuppressive Pathway Unleash Immune Responses in Combination Cancer Therapies. Cell Rep. *27*, 2411–2425.e9.

Petralia, F., Wang, P., Yang, J., and Tu, Z. (2015). Integrative random forest for gene regulatory network inference. Bioinformatics *31*, i197–i205.

Petralia, F., Song, W.M., Tu, Z., and Wang, P. (2016). New Method for Joint Network Analysis Reveals Common and Different Coexpression Patterns among Genes and Proteins in Breast Cancer. J. Proteome Res. *15*, 743–754.

Petralia, F., Wang, L., Peng, J., Yan, A., Zhu, J., and Wang, P. (2018). A new method for constructing tumor specific gene co-expression networks based on samples with tumor purity heterogeneity. Bioinformatics *34*, i528–i536.

Pilarczyk, M., Najafabadi, M.F., Kouril, M., Vasiliauskas, J., Niu, W., Shamsaei, B., Mahi, N., Zhang, L., Clark, N., Ren, Y., et al. (2019). Connecting omics signatures of diseases, drugs, and mechanisms of actions with iLINCS. bioRxiv. https://doi.org/10.1101/826271.

Porro, F., Rosato-Siri, M., Leone, E., Costessi, L., Iaconcig, A., Tongiorgi, E., and Muro, A.F. (2010). beta-adducin (Add2) KO mice show synaptic plasticity, motor coordination and behavioral deficits accompanied by changes in the expression and phosphorylation levels of the alpha- and gamma-adducin subunits. Genes Brain Behav. *9*, 84–96.

Pratilas, C.A., Taylor, B.S., Ye, Q., Viale, A., Sander, C., Solit, D.B., and Rosen, N. (2009). (V600E)BRAF is associated with disabled feedback inhibition of RAF-MEK signaling and elevated transcriptional output of the pathway. Proc. Natl. Acad. Sci. USA *106*, 4519–4524.

Pugh, T.J., Weeraratne, S.D., Archer, T.C., Pomeranz Krummel, D.A., Auclair, D., Bochicchio, J., Carneiro, M.O., Carter, S.L., Cibulskis, K., Erlich, R.L., et al. (2012). Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. Nature *488*, 106–110.

Qiao, S., Peng, R., Yan, H., Gao, Y., Wang, C., Wang, S., Zou, Y., Xu, X., Zhao, L., Dong, J., et al. (2014). Reduction of phosphorylated synapsin I (ser-553) leads to spatial memory impairment by attenuating GABA release after microwave exposure in Wistar rats. PLoS ONE *9*, e95503.

Rivero-Hinojosa, S., Lau, L.S., Stampar, M., Staal, J., Zhang, H., Gordish-Dressman, H., Northcott, P.A., Pfister, S.M., Taylor, M.D., Brown, K.J., and Rood, B.R. (2018). Proteomic analysis of Medulloblastoma reveals functional biology with translational potential. Acta Neuropathol. Commun. *6*, 48.

Robinson, G., Parker, M., Kranenburg, T.A., Lu, C., Chen, X., Ding, L., Phoenix, T.N., Hedlund, E., Wei, L., Zhu, X., et al. (2012). Novel mutations target distinct subgroups of medulloblastoma. Nature *488*, 43–48.

Salomonis, N., Dexheimer, P.J., Omberg, L., Schroll, R., Bush, S., Huo, J., Schriml, L., Ho Sui, S., Keddache, M., Mayhew, C., et al. (2016). Integrated Genomic Analysis of Diverse Induced Pluripotent Stem Cells from the Progenitor Cell Biology Consortium. Stem Cell Reports *7*, 110–125.

Santamaría, D., Barrière, C., Cerqueira, A., Hunt, S., Tardy, C., Newton, K., Cáceres, J.F., Dubus, P., Malumbres, M., and Barbacid, M. (2007). Cdk1 is sufficient to drive the mammalian cell cycle. Nature *448*, 811–815.

Savage, S.R., Shi, Z., Liao, Y., and Zhang, B. (2019). Graph Algorithms for Condensing and Consolidating Gene Set Analysis Results. Mol. Cell. Proteomics *18* (8, suppl 1), S141–S152.

Schmitt, J.M., and Stork, P.J. (2001). Cyclic AMP-mediated inhibition of cell growth requires the small G protein Rap1. Mol. Cell. Biol. *21*, 3671–3683.

Schreck, K.C., Grossman, S.A., and Pratilas, C.A. (2019). BRAF Mutations and the Utility of RAF and MEK Inhibitors in Primary Brain Tumors. Cancers (Basel) *11*, 1262.

Shah, K., and Lahiri, D.K. (2017). A Tale of the Good and Bad: Remodeling of the Microtubule Network in the Brain by Cdk5. Mol. Neurobiol. *54*, 2255–2268.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Shin, H.J., Lee, S., and Jung, H.J. (2019). A curcumin derivative hydrazinobenzoylcurcumin suppresses stem-like features of glioblastoma cells by targeting Ca^{2+} /calmodulin-dependent protein kinase II. J. Cell. Biochem. *120*, 6741–6752.

Shteynberg, D.D., Deutsch, E.W., Campbell, D.S., Hoopmann, M.R., Kusebauch, U., Lee, D., Mendoza, L., Midha, M.K., Sun, Z., Whetton, A.D., and Moritz, R.L. (2019). PTMProphet: Fast and Accurate Mass Modification Localization for the Trans-Proteomic Pipeline. J. Proteome Res. *18*, 4262–4272.

Sokolov, A., Paull, E.O., and Stuart, J.M. (2016). One-Class Detection of Cell States in Tumor Subtypes. Pac. Symp. Biocomput. *21*, 405–416.

Solga, A.C., Pong, W.W., Walker, J., Wylie, T., Magrini, V., Apicelli, A.J., Griffith, M., Griffith, O.L., Kohsaka, S., Wu, G.F., et al. (2015). RNA-sequencing reveals oligodendrocyte and neuronal transcripts in microglia relevant to central nervous system disease. Glia *63*, 531–548.

Song, X., Ji, J., Gleason, K.J., Yang, F., Martignetti, J.A., Chen, L.S., and Wang, P. (2019). Insights into Impact of DNA Copy Number Alteration and Methylation on the Proteogenomic Landscape of Human Ovarian Cancer via a Multi-omics Integrative Analysis. Mol. Cell. Proteomics *18* (8, suppl 1), S52–S65.

Staedtler, F., Hartmann, N., Letzkus, M., Bongiovanni, S., Scherer, A., Marc, P., Johnson, K.J., and Schumacher, M.M. (2013). Robust and tissue-independent gender-specific transcript biomarkers. Biomarkers *18*, 436–445.

Stathias, V., Turner, J., Koleti, A., Vidovic, D., Cooper, D., Fazel-Najafabadi, M., Pilarczyk, M., Terryn, R., Chung, C., Umeano, A., et al. (2020). LINCS Data Portal 2.0: next generation access point for perturbation-response signatures. Nucleic Acids Res. *48* (D1), D431–D439.

Stenson, P.D., Mort, M., Ball, E.V., Evans, K., Hayden, M., Heywood, S., Hussain, M., Phillips, A.D., and Cooper, D.N. (2017). The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum. Genet. *136*, 665–677.

Stepulak, A., Rola, R., Polberg, K., and Ikonomidou, C. (2014). Glutamate and its receptors in cancer. J. Neural Transm. *121*, 933–944.

Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell *171*, 1437–1452.e17.

Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput. Biol. *12*, e1004873.

Tanaka, H., Sasayama, T., Tanaka, K., Nakamizo, S., Nishihara, M., Mizukawa, K., Kohta, M., Koyama, J., Miyake, S., Taniguchi, M., et al. (2013). MicroRNA-183 upregulates HIF-1α by targeting isocitrate dehydrogenase 2 (IDH2) in glioma cells. J. Neurooncol. *111*, 273–283.

Tomson, B.N., and Arndt, K.M. (2013). The many roles of the conserved eukaryotic Paf1 complex in regulating transcription, histone modifications, and disease states. Biochim. Biophys. Acta *1829*, 116–126.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA *98*, 5116–5121.

Uchida, S., Martel, G., Pavlowsky, A., Takizawa, S., Hevi, C., Watanabe, Y., Kandel, E.R., Alarcon, J.M., and Shumyatsky, G.P. (2014). Learning-induced and stathmin-dependent changes in microtubule stability are critical for memory and disrupted in ageing. Nat. Commun. *5*, 4389.

Venkataramani, V., Tanev, D.I., Strahle, C., Studier-Fischer, A., Fankhauser, L., Kessler, T., Körber, C., Kardorff, M., Ratliff, M., Xie, R., et al. (2019). Glutamatergic synaptic input to glioma cells drives brain tumour progression. Nature *573*, 532–538.

Venkatesh, H.S., Morishita, W., Geraghty, A.C., Silverbush, D., Gillespie, S.M., Arzt, M., Tam, L.T., Espenel, C., Ponnuswami, A., Ni, L., et al. (2019). Electrical and synaptic integration of glioma into neural circuits. Nature *573*, 539–545.

Vu, N.T., Park, M.A., Shultz, J.C., Goehe, R.W., Hoeferlin, L.A., Shultz, M.D., Smith, S.A., Lynch, K.W., and Chalfant, C.E. (2013). hnRNP U enhances caspase-9 splicing and is modulated by AKT-dependent phosphorylation of hnRNP L. J. Biol. Chem. *288*, 8575–8584.

Wang, X., Park, J., Susztak, K., Zhang, N.R., and Li, M. (2019). Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat. Commun. *10*, 380.

Whiteaker, J.R., Zhao, L., Saul, R., Kaczmarczyk, J.A., Schoenherr, R.M., Moore, H.D., Jones-Weinert, C., Ivey, R.G., Lin, C., Hiltke, T., et al. (2018). A Multiplexed Mass Spectrometry-Based Assay for Robust Quantification of Phosphosignaling in Response to DNA Damage. Radiat. Res. *189*, 505–518.

Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572–1573.

Yoo, S., Huang, T., Campbell, J.D., Lee, E., Tu, Z., Geraci, M.W., Powell, C.A., Schadt, E.E., Spira, A., and Zhu, J. (2014). MODMatcher: multi-omics data matcher for integrative genomic analysis. PLoS Comput. Biol. *10*, e1003790.

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun. *4*, 2612.

Yuan, J., Levitin, H.M., Frattini, V., Bush, E.C., Boyett, D.M., Samanamud, J., Ceccarelli, M., Dovas, A., Zanazzi, G., Canoll, P., et al. (2018). Single-cell transcriptome analysis of lineage diversity in high-grade glioma. Genome Med. *10*, 57.

Zecha, J., Satpathy, S., Kanashova, T., Avanessian, S.C., Kane, M.H., Clauser, K.R., Mertins, P., Carr, S.A., and Kuster, B. (2019). TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach. Mol. Cell. Proteomics *18*, 1468–1478.

Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., Chambers, M.C., Zimmerman, L.J., Shaddox, K.F., Kim, S., et al.; NCI CPTAC (2014). Proteogenomic characterization of human colon and rectal cancer. Nature *513*, 382–387.

Zhang, H., Liu, T., Zhang, Z., Payne, S.H., Zhang, B., McDermott, J.E., Zhou, J.Y., Petyuk, V.A., Chen, L., Ray, D., et al.; CPTAC Investigators (2016). Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. Cell *166*, 755–765.

Zhang, Y., Lv, W., Li, Q., Wang, Q., Ru, Y., Xiong, X., Yan, F., Pan, T., Lin, W., and Li, X. (2019). IDH2 compensates for IDH1 mutation to maintain cell survival under hypoxic conditions in IDH1-mutant tumor cells. Mol. Med. Rep. *20*, 1893–1900.

Zhao, L., Whiteaker, J.R., Pope, M.E., Kuhn, E., Jackson, A., Anderson, N.L., Pearson, T.W., Carr, S.A., and Paulovich, A.G. (2011). Quantification of proteins using peptide immunoaffinity enrichment coupled with mass spectrometry. J. Vis. Exp. (53), 2812.

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit monoclonal anti-TP53	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-TP53-1; RRID: AB_2877664
Rabbit monoclonal anti-CDK1	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-CDK1-1; RRID: AB_2877655
Rabbit monoclonal anti-TUBB	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-TUBB-1; RRID: AB_2877665
Rabbit monoclonal anti-RAF1	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-RAF1-3; RRID: AB_2827856
Rabbit monoclonal anti-KRT7	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-KRT7-1; RRID: AB_2820264
Rabbit monoclonal anti-KRT7	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-KRT7-1; RRID: AB_2820264
Rabbit monoclonal anti-KRT5	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-KRT5-1; RRID: AB_2820262
Rabbit monoclonal anti-KRT5	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-KRT5-2; RRID: AB_2820263
Rabbit monoclonal anti-FBN1	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-FBN1-1; RRID: AB_2877658
Rabbit monoclonal anti-FBN1	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-FBN1-2; RRID: AB_2877659
Rabbit monoclonal anti-PCNA	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-PCNA-2; RRID: AB_2877661
Rabbit monoclonal anti-PCNA	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-PCNA-2; RRID: AB_2877661
Rabbit monoclonal anti-ATM	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-ATM-3; RRID: AB_2877654
Rabbit monoclonal anti-ATM	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-ATM-1; RRID: AB_2877653
Rabbit monoclonal anti-FAAP100	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-FAAP100-1; RRID: AB_2877656
Rabbit monoclonal anti-FAAP100	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-FAAP100-2; RRID: AB_2877657
Rabbit monoclonal anti-RIF1	Fred Hutchinson Cancer Research Center	custom monoclonal; RRID: AB_2877666
Mouse monoclonal anti-PTEN	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-PTEN-4; RRID: AB_2722096
Mouse monoclonal anti-PTEN	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-PTEN-7; RRID: AB_2722099
Mouse monoclonal anti-CCND1	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-CCND1-5; RRID: AB_2722035
Mouse monoclonal anti-RPTOR	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-RPTOR-2; RRID: AB_2722104
Mouse monoclonal anti-RPTOR	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-RPTOR-1; RRID: AB_2722103
Mouse monoclonal anti-FOS	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-FOS-2; RRID: AB_2722059
Mouse monoclonal anti-FOS	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-FOS-3; RRID: AB_2722060

(*Continued on next page*)

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse monoclonal anti-AKT2	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-AKT2-3; RRID: AB_2722017
Mouse monoclonal anti-CDH1	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-CDH1-1; RRID: AB_2722037
Mouse monoclonal anti-CDH1	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-CDH1-2; RRID: AB_2722038
Mouse monoclonal anti-BRAF	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-BRAF-1; RRID: AB_2722024
Mouse monoclonal anti-RAF1	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-RAF1-1; RRID: AB_2877663
Mouse monoclonal anti-MTOR	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-MTOR-5; RRID: AB_2722089
Mouse monoclonal anti-MTOR	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-MTOR-4; RRID: AB_2722088
Mouse monoclonal anti-MTOR	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-MTOR-3; RRID: AB_2722087
Mouse monoclonal anti-MTOR	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-MTOR-7; RRID: AB_2722091
Mouse monoclonal anti-ERBB2	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-ERBB2-3; RRID: AB_2722053
Mouse monoclonal anti-ERBB2	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-ERBB2-2; RRID: AB_2722052
Mouse monoclonal anti-AKT1	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-AKT1-1; RRID: AB_2722012
Mouse monoclonal anti-AKT1	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-AKT1-3; RRID: AB_2722014
Mouse monoclonal anti-AKT3	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-AKT3-6; RRID: AB_2814779
Rabbit monoclonal anti-EGFR	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-EGFR-8; RRID: AB_2827847
Rabbit monoclonal anti-EGFR	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-EGFR-9; RRID: AB_2827848
Rabbit monoclonal anti-GSK3B	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-GSK3B-10; RRID: AB_2827872
Rabbit monoclonal anti-GSK3B	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-GSK3B-9; RRID: AB_2827863
Rabbit monoclonal anti-MAPK1	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-MAPK1-2; RRID: AB_2827853
Rabbit monoclonal anti-MAPK3	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-MAPK1-2; RRID: AB_2827853
Rabbit monoclonal anti-MTOR	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-MTOR-9; RRID: AB_2820270
Rabbit monoclonal anti-MAPK1	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-MAPK3-4; RRID: AB_2827841
Rabbit monoclonal anti-MAPK3	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-MAPK3-4; RRID: AB_2827841
Rabbit monoclonal anti-RIF1	Fred Hutchinson Cancer Research Center	custom monoclonal; RRID: AB_2877666
Rabbit monoclonal anti-ERBB2	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-ERBB2-1; RRID: AB_2617255
Rabbit monoclonal anti-JUN	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-JUN-4; RRID: AB_2877660
Rabbit monoclonal anti-JUN	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-JUN-4; RRID: AB_2877660

(*Continued on next page*)

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Rabbit monoclonal anti-RB1	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-RB1-2; RRID: AB_2820275
Rabbit monoclonal anti-ANXA1	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-ANXA1-4; RRID: AB_2617223
Rabbit monoclonal anti-CALR	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-CALR-1; RRID: AB_2617228
Rabbit monoclonal anti-GSTP1	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-GSTP1-1; RRID: AB_2617266
Rabbit monoclonal anti-PCNA	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-PCNA-1; RRID: AB_2617307
Rabbit monoclonal anti-RRM2	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-RRM2-1; RRID: AB_2617333
Rabbit monoclonal anti-FSCN1	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-FSCN1-1; RRID: AB_1553796
Rabbit monoclonal anti-PRDX4	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-PRDX4-5; RRID: AB_2877662
Rabbit monoclonal anti-UBE2C	CPTAC Antibody Portal (https://antibodies.cancer.gov/)	CPTC-UBE2C-2; RRID: AB_2617367
Biological Samples		
Fresh frozen tissue samples	See STAR Methods	N/A
Chemicals, Peptides, and Recombinant Proteins		
Papain	SciQuest	Catalog: LS003124
Ovomucoid	SciQuest	Catalog: 542000
DNase	Sigma	Catalog: 10104159001
RNase A	QIAGEN	Catalog: 19101
TMT 11plex reagents	Thermo Fisher Scientific	Catalog: A34808
IMAC resin kit	Thermo Fisher Scientific	Catalog: A32992
BCA protein assay	Thermo Fisher Scientific	Catalog: 23221
Stable isotope-labeled synthetic peptide standards	New England Peptide	Custom synthesis, > 95% pure by HPLC
Aprotinin	Sigma	Catalog: A6103
Leupeptin	Roche	Catalog: 11017101001
Phenylmethylsulfonyl fluoride	Sigma	Catalog: 93482
Phosphatase Inhibitor Cocktail 2	Sigma	Catalog: P5726
Phosphatase Inhibitor Cocktail 3	Sigma	Catalog: P0044
PUGNAc	Sigma	Catalog: A7229
Sodium butyrate	Sigma	Catalog: 303410
Ni-NTA Superflow Agarose Beads	QIAGEN	Catalog: 30410
Lysyl Endopeptidase	Wako Chemicals	Catalog 129-02541
Sequencing grade modified trypsin	Promega	Catalog: V517
Critical Commercial Assays		
TruSeq RNA Sample Prep Kit	Illumina	Catalog: FC-122-1001
KAPA Library Preparation Kit	Roche	Catalog: KK8201
DNA/RNA AllPrep Kit	QIAGEN	Catalog: 80204
AllPrep DNA/RNA/miRNA Universal kit	QIAGEN	Catalog: 80224
QIAsymphony DSP DNA Midi Kit	QIAGEN	Catalog: 937255
KAPA Hyper prep kit	Roche	Catalog: KK8541
KAPA Stranded RNA-Seq with RiboErase kit	Roche	Catalog: KK8484

(*Continued on next page*)

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
PhosphoSitePlus	Hornbeck et al., 2015	https://www.phosphosite.org
TCGA – GBM	Cancer Genome Atlas Research Network, 2008	https://portal.gdc.cancer.gov/
TCGA – LGG	Brat et al., 2015	https://portal.gdc.cancer.gov/
Software and Algorithms		
Philosopher	Alexey Nesvizhskii Lab	https://philosopher.nesvilab.org/
PeptideProphet	Keller et al., 2002	http://peptideprophet.sourceforge.net/
ComBat (v3.20.0)	Johnson et al., 2007	https://bioconductor.org/packages/release/bioc/html/sva.html
DreamAI	Ma et. al., 2020	https://github.com/WangLab-MSSM/DreamAI
GISTIC2.0	Mermel et al., 2011	ftp://ftp.broadinstitute.org/pub/GISTIC2.0/GISTIC_2_0_23.tar.gz
iProFun	Song et al., 2019	https://github.com/WangLab-MSSM/iProFun
ESTIMATE	Yoshihara et al., 2013	https://bioinformatics.mdanderson.org/public-software/estimate/
Joint Random Forest	Petralia et al., 2016	https://rdrr.io/cran/JRF/man/JRF.html
TSNet	Petralia et al., 2018	https://github.com/WangLab-MSSM/TSNet
xCell	Aran et al., 2017	https://xcell.ucsf.edu/
CPTAC Data Viewer	Pei Wang Lab	http://pbt.cptac-data-view.org/
iCAVE	Liluashvili et al., 2017	https://labs.icahn.mssm.edu/gumuslab/software
ConsensusClusterPlus	Monti et al., 2003; Wilkerson and Hayes, 2010	http://bioconductor.org/packages/release/bioc/html/CancerSubtypes.html
Strelka2 v2.9.3	Kim et al., 2018b	https://github.com/Illumina/strelka
CNVkit v. 2.9.3	Talevich et al., 2016	https://github.com/etal/cnvkit
STAR v2.6.1d	Dobin et al., 2013	https://github.com/alexdobin/STAR
GENCODE v27	GENCODE consortium	https://www.gencodegenes.org/human/release_27.html
RSEM v1.3.1	Li and Dewey, 2011	https://github.com/deweylab/RSEM
Sumer	Savage et al., 2019	https://github.com/bzhanglab/sumer
TCGAbiolinks	Colaprico et al., 2016	https://bioconductor.org/packages/release/bioc/html/TCGAbiolinks.html
MoonlightR	Colaprico et al., 2020	https://bioconductor.org/packages/release/bioc/html/MoonlightR.html
Music	Wang et al., 2019	https://github.com/xuranw/MuSiC
Ascore v1.0.6858	Beausoleil et al., 2006	https://github.com/PNNL-Comp-Mass-Spec/AScore
MASIC	Monroe et al., 2008	https://github.com/PNNL-Comp-Mass-Spec/MASIC
MS-GF+ v9981	Kim and Pevzner, 2014	https://github.com/MSGFPlus/msgfplus
mzRefinery	(Gibbons et al., 2015)	https://omics.pnl.gov/software/mzrefinery

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Pei Wang (pei.wang@mssm.edu).

Materials Availability

N/A

Data and Code Availability

All raw genomic data is available upon access request through the Children's Brain Tumor Network (www.cbtn.org) and can be accessed through the Gabriella Miller Kids First Portal (https://kidsfirstdrc.org/). Processed genomic data is available through the Open Pediatric Brain Tumor Atlas (https://github.com/AlexsLemonade/OpenPBTA-analysis). All raw proteomics data and processed proteogenomic data are available through the Clinical Proteomic Tumor Analysis Consortium Data Portal (https://cptac-data-portal.georgetown.edu/cptacPublic/) and the Proteomics Data Commons (https://pdc.cancer.gov/pdc/). In addition, all processed proteogenomic datasets as well as clinical meta information can be queried, visualized and downloaded from an interactive *ProTrack* data portal (http://pbt.cptac-data-view.org/), as well as through the *PedcBioPortal* (https://pedcbioportal.kidsfirstdrc.org:443/saml/discovery?entityID=d3b-center.auth0.com&returnIDParam=idp).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patient selection for the discovery cohort

The samples were obtained from the Children's Brain Tumor Network (CBTN) at the Children's Hospital of Philadelphia (CHOP). The patient selection was built based on specimen availability and defined two broad classes of tumors: (1) High grade tumors driven by epigenetic dysregulation (HGG, DIPG, ATRT and/or other embryonal tumors) and (2) Low grade tumors defined by receptor-tyrosine kinase and MAPK signaling alterations including kinase fusions. Additional associated clinical determinants for cohort selection included: (A) Tumor histologies for which there is more than one therapeutic standard of care and for which a multidimensional proteogenomic analysis could further inform an assessment of therapeutic response; (B) Tumor histologies for which genomic alterations and/or classification have failed to provide differential prognosis; (C) Tumor cohorts for which comprehensive profiling could inform the course of metastasis. These considerations led to the selection of 226 samples from 204 pediatric subjects treated surgically and clinically at the Children's Hospital of Philadelphia for whom deep longitudinal, clinical data is available.

Sample collection for the discovery cohort

Samples were collected at the time of surgery (217 samples) or autopsy (9 samples), flash-frozen, and stored in BioRC (Biorepository Resource Center) at Children's Hospital of Philadelphia. Frozen tissue pieces ~75mg were cut off using disposable scalpels on dry ice and delivered to Fred Hutchinson Cancer Research Center for sample preparation for proteomics profiling. ~20 mg frozen tissue and up to 0.4-1ml of blood was used for nucleic acid extractions, which were performed at the Biorepository Resource Center at Children's Hospital of Philadelphia.

Sample collection for the HGG validation study

For the validation studies, the specimens from 41 patient subjects were collected through Children's Brain Tumor Network (CBTN) sites including Children's Hospital of Philadelphia (CHOP), Seattle Children's Hospital, Meyer Children's Hospital, UCSF Benioff Children's Hospital, University of Pittsburgh, Lurie Children's Hospital, Children's National Medical Center) and through the HUP-CHOP Neurosurgery Tumor Tissue Bank Collaborative at the Hospital of University of Pennsylvania. Among the 41, 18 were part of the discovery cohort who had remaining tumor materials. All samples were fresh frozen collected at the time of surgery, shipped and stored in BioRC (Biorepository Resource Center) at Children's Hospital of Philadelphia. ~30mg tissue pieces were cut/chipped off using disposable scalpels on dry ice and delivered to Fred Hutchinson Cancer Research Center for sample preparation for proteomics profiling.

METHOD DETAILS

Nucleic acid extractions, WGS and RNaseq

Tissues were lysed with QIAGEN TissueLyser II (QIAGEN) using 5 mm steel beads (cat# 69989, QIAGEN) 2 × 30 s at 18Hz settings, and processed with CHCl3 extraction and run on the QiaCube automated platform (QIAGEN) using the AllPrep DNA/RNA/miRNA Universal kit (cat# 80224, QIAGEN). Thawed blood was RNase A (cat#, 19101, QIAGEN) treated and processed using the QIAGEN QIAsymphony automated platform (QIAGEN) using the QIAsymphony DSP DNA Midi Kit (cat# 937255, QIAGEN). DNA and RNA quantity and quality was assessed by PerkinElmer DropletQuant UV-VIS spectrophotometer (PerkinElmer) and an Agilent 4200 TapeStation (Agilent, USA) for RINe and DINe (RNA Integrity Number equivalent and DNA Integrity Number equivalent respectively). Library preparation and sequencing was performed by the NantHealth sequencing center. Briefly, DNA sequencing libraries were prepared for both tumor tissue and matched-germline (blood) DNA using the KAPA Hyper prep kit (cat# KK8541, Roche); Whole genome sequencing (WGS) was performed at an average coverage of 60X for tumor samples and 30X for matched-germline. The panel tumor sample was sequenced to 470X and the normal panel sample was sequenced to 308X. Tumor RNA-Seq libraries were prepared using KAPA Stranded RNA-Seq with RiboErase kit (cat# KK8484, Roche). RNA samples were sequenced to an average of 200M reads. All sequencing was performed on the Illumina HiSeq platform (X/400) (Illumina) with 2 × 150bp read length.

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

Somatic Mutation and CNV calling

Strelka2 (Kim et al., 2018b) v2.9.3 was run for canonical chromosomes (chr1-22, X,Y,M) using default parameters and the resulting VCF was filtered for PASS variants. Gene level mutation status were summarized based on somatic mutations detected in coding regions, having minimum sequencing depth of 30, and minimum alternative variant count of 5.

CNVkit v. 2.9.3 (Talevich et al., 2016) was run in batch wgs mode, paired tumor-normal, using the hg38 annotation reference from UCSC (http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/refFlat.txt.gz). All output files, such as seg, gain/loss, and scatter/diagram plots were generated using CNVkit's export and other built-in functions.

RNaseq data preprocessing

STAR v2.6.1d (Dobin et al., 2013) was used to align paired-end RNA-seq reads against the ENSEMBL GENCODE 27 "Comprehensive gene annotation" reference (https://www.gencodegenes.org/human/release_27.html). RSEM v1.3.1 (Li and Dewey, 2011) was used to generate both FPKM and TPM transcript- and gene-level expression values. Then, log2(x+1) transform was applied and samples with replicates were averaged.

Proteomic experiments for Discovery Cohort

11-Plex Preparation

Sample preparation for MS analysis was performed as described previously (Navarrete-Perea et al., 2018). Lysates were prepared from 226 Cryo-pulverized human Pediatric Brain Tumor samples in lysis buffer (6M Urea, 25 mM Tris, pH8.0, 1 mM EDTA, 1 mM EGTA, Sigma Protease Inhibitor Cat# P8340, Sigma Phosphatase Cocktail Cat# P5726, Sigma Phosphatase Cocktail Cat# P0044) supplied by Fred Hutch. Lysates were reduced with 5 mM neutralized TCEP (Pierce, #77720) for 15 min., alkylated with 10 mM Iodoacetamide (Sigma, #A3221) for 30 minutes in the dark and quenched with 5 mM Dithiothreitol (Thermo Scientific, #20291) for 15 mins. Protein was precipitated with methanol-chloroform, and the protein pellet was resuspended in 200 mM EPPS (pH 8.0). The samples were digested sequentially with Lys-C protease (Wako, 129-02541, 2 mg/mL Stock) at a 100:1 protein-to-protease ratio with constant shaking overnight at room temperature followed by Trypsin (Pierce, 90305, 1 mg/mL stock) digestion at a 100:1 protein-to-protease ratio for another 6 h at 37°C. Digested peptides were assayed for peptide concentration with Pierce Quantitative Colorimetric Peptide Assay (#23275) as per Manufacturer's protocol.

11-plex Experimental Layout

Proteome and Phosphoproteome analysis of pediatric brain cancer samples were structured as TMT11-plex experiments. 226 unique samples plus a few replicates and QC samples were arranged in twenty-three 11-plex experiments with 10 individual samples occupying the first 10 channels of each experiment and the 11th channel being "Bridge Channel" i.e., Common Reference Sample, used for quantitative comparison across all sample sets. To prepare the bridge channel which broadly represents the population of pediatric brain cancer samples in our experiment, digested peptides from indicated samples were pooled together.

TMT Labeling of Peptides and Quality Check

About 100 μg of digested peptides per sample were labeled with TMT11-plex reagent according to the manufacturer's instructions (Thermo Scientific, Pierce Biotechnology, Germany). About 2 μg of each sample from each 11-plex experiment was removed and combined in 100 μl of 1% formic acid (FA) for a quality control check. The remaining samples were frozen immediately at −80°C for future quenching and HPLC fractionation. The combined samples in 100 μl of 1% FA from each 11-plex experiment were desalted by StageTip containing 4 small (0.9 mm) discs of 3M Empore C18 material following standard procedure. Eluted material was dried by speedvac, resuspended in 5% ACN/5% FA and analyzed by a mass spectrometer to check (a) digestion efficiency, (b) labeling efficiency & (c) summed signal-to-noise ratios among samples. As a standard of quality check (QC), minimum of 97% labeled MS/MS spectra, a maximum of 5% missed cleavage rate, and summed signal-to-noise ratio variations of < 1.5-fold within each plex were required to proceed further. Following successful QC checks, unwanted TMT labeling of tyrosine residues was reversed with a final concentration of ~0.3% (v/v) hydroxylamine (Sigma, 467804, 50% stock) for 15 mins and finally quenched with 50% TFA to a final concentration of ~0.5% (v/v). Labeled peptides from each of the twenty three 11-plex experiments were combined into 23 samples, acidified, and subsequently desalted on C18 Sep-Pak columns. Eluates were dried by SpeedVac in preparation for Phosphopeptide Enrichment.

Phosphopeptide Enrichment

High-Select Fe-NTA Phosphopeptide Enrichment Kits (Thermo Scientific; #A32992) were used for phosphopeptide enrichment step ("mini-phos") as per protocol described (Navarrete-Perea et al., 2018). Lyophilized labeled peptide sample was completely dissolved in 200 μL of Binding/Wash buffer with vortexing, and pH was confirmed to be below 3. Samples were loaded onto spin columns equilibrated per Manufacturer's method and mixed by gentle tapping until the resin was in suspension. Samples were incubated for 30 minutes at room temperature with gentle mixing every 5 minutes. Following incubation, the spin columns were placed in a microfuge tube, centrifuged at 1000 × g for 30 s, and washed thrice with a Binding/Wash buffer. All flow-through fractions were collected in the same tube, desalted, dried, resuspended and directed to basic-pH HPLC for global proteome analysis. Phosphopeptide-bound spin columns were placed in a new microfuge tube, containing 1% FA and Phosphopeptides were eluted with an Elution buffer and dried immediately by speedvac. For phosphoproteome analysis, phosphopeptide enriched samples were resuspended in 1% FA, desalted by stage tip and eluted into Agilent deactivated glass vial inserts, dried by speedvac and finally phosphopeptide enriched samples were resuspended with 10 μL of 5% FA and made ready to be analyzed by LC-MS/MS analysis.

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

Offline fractionation of peptides

To reduce sample complexity, peptide samples were separated by high pH reversed phase (RP) fractionation. Phosphopeptide Flow-Through fractions were checked for pH (pH < 3) and desalted on C18 Sep-Pak columns. Eluates were dried in a Speed Vacuum Concentrator, reconstituted in 500 μL of Buffer A (10 mM ammonium bicarbonate, 5% Acetonitrile, pH 8), loaded onto an Agilent 300Extend C18 column (3.5 μm bead size, 4.6 mm ID and 220 mm long), and separated on an Agilent 1200 HPLC instrument at a flow rate of 0.6 mL/min with a 60 min linear gradient from 13% to 42% buffer B (10 mM ammonium bicarbonate, 90% ACN, pH 8) into a total of 96 fractions. Each fraction contained ~500uL, at ~37 s per fraction. All 96 fractions were consolidated into 12 final fractions by column, desalted with stage-tip, resuspended with 10 μL of 5% ACN - 5% FA and made ready to be analyzed by LC-MS/MS analysis.

Mass Spectrometry (MS) Instrument

Global proteome analyses were performed on an Orbitrap Fusion Tribrid Mass Spectrometer and phosphoproteome analyses were performed on an Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Scientific) both in-line with a NanoSpray Flex NG ion source using MS3-based TMT centric mass spectrometer method. Both the instruments are online with a Liquid Chromatography System (EASY-nLC 1200 System).

Online Liquid Chromatography

Online separation was performed on a nano-flow UHPLC EASY-nLC 1200 system (Thermo Scientific). In this set up, the LC system with a column in (IDxL 20 mm x 550 mm; LC 560) from sample valve and a column out (IDxL 75 mm x 550 mm; LC 562) to waste valve, a Fused Silica Capillary Tubing, that delivers sample to MS were connected via a stainless-steel cross union (Micro-Cross Assay-SS 360 μm UH-906/ Western Analytical products). A platinum wire was used to deliver electrospray source voltage. The column was heated to 60°C using a column heater sleeve (Phoenix-ST) to prevent over-pressuring of columns during UHPLC separation. The capillary tubing (Inner Diameter: 100μm, Outer Diameter: 375μm) was pulled to an opening of 10 mm and packed in-house with 2.6 mm beads (90 A Pore diameter, Thermo Scientific) slurry made in Buffer B (90% Acetonitrile, 0.1% FA). Each analysis, ~1 mg of peptide in a 1-5 mL injection volume based on the sample dilution, and for each Phosphoproteome sample in a 5 ml, were loaded onto the column in Mobile phase, comprised of 0.1% FA (Buffer A). LC-MS/MS method consisted of an initial 10 min column-equilibration procedure and a 20 min sample-loading procedure, both at 800 bar. For global proteome analyses, the peptides were separated using a 150 min gradient of 5 to 42% Acetonitrile in 0.1% FA and respective flow rate was adjusted to separate the fraction within a pressure difference of 300 – 400 bar. Phosphopeptides were separated over 160 min with gradient of 3 to 30% Acetonitrile in 0.1% FA.

Mass Spectrometry Analyses

For data-dependent experiments (MS2 and MS3), all instrument operational parameters were specified through the instrument method editor. Data-dependent acquisition was performed using Xcalibur v2.1 software in positive ion mode at a spray voltage of 2.6 kV and 300oC ion transfer tube temperature. For 276 Peptide fractions analyses, MS1 Spectra was detected with Orbitrap at a resolution of 120K, scan range (m/z) being 350 – 1350 and AGC target being 1.0e6 with 50 ms maximum ion injection time. For MS2 analysis, top ten precursors were selected with peptide as monoisotopic peak determination, intensity threshold of 1.0e3, and charge state screening was enabled to include only precursor charge states 2-6. Peptides that triggered MS/MS scans were dynamically excluded from further MS/MS scans for 90 s, with a ± 10 ppm mass tolerance. Perform dependent scan on single charge state per precursor only and Exclude within the cycle were enabled. In data-dependent charge specific MS2 analysis, ions were first isolated by Quadruple with an isolation window of 0.7 or 0.5 (based on instruments used) and activated at ion Trap with CID collision Energy being 35% in 10 ms and activation Q of 0.25. Ion trap detection were set to normal scan range mode with rapid ion trap scan rate, 9.0e3 AGC target and 80 ms ion injection time. Following the acquisition of each MS2 fragment ion, precursors were selected with a mass range (m/z) between 400 – 2000 with a mass exclusion width 50 (low) and 5 (high). About 10 precursor fragment ions were simultaneously isolated by SPS selection at a time for every MS3 precursor population, which then fragmented by HCD with HCD collision energy of 55% and fragmented reporter ions with a normal scan range mode were analyzed in Orbitrap at a resolution of 50K, AGC Target 1.0e5 and 120 ms maximum ion injection time. To further minimize the influence of co-eluting species, peptides with isolation specificities less than 1.2 for + 2 charge, 1.0 for + 3 charge and 0.8 for 4 - 6 charge.

Each of the 23 Phosphopeptide enriched samples were directed to both CID and High Resolution HCD activation using similar multinotch MS3-based TMT methods. For CID activation, both multistage activation and injections for all available parallelization time were enabled with ion trap scan rate set to Turbo and neutral loss mass to 97.9763. For HCD activation during MS2 analysis, HCD collision energy was set to 32%. Both CID and HCD activation are considered as two fractions from each of the 23 Phosphopeptides enriched samples, totaling 46 Peptide fractions.

Protein Identification and Quantification

Proteomics processing of whole proteome and phosphopeptide-enriched datasets was performed as described previously (Clark et al., 2019; Djomehri et al., 2020). MSFragger version 20190628 (Kong et al., 2017) was used to search a CPTAC harmonized RefSeq protein sequence database appended with an equal number of decoy sequences. Specifically, the RefSeq-based sequence database containing 41,457 proteins mapped to the human reference genome (GRCh38/hg38) obtained via the UCSC Table Browser (https://genome.ucsc.edu/cgi-bin/hgTables) on June 29, 2018, with the addition of 13 proteins encoded in the human mitochondrial genome, and 264 common laboratory contaminant proteins. MS/MS spectra were searched using a precursor-ion mass tolerance of

20 ppm, fragment mass tolerance of 0.7 Da, and allowing C12/C13 isotope errors (−1/0/1/2/3 for global, and 0/1/2/3 for phosphopeptide-enriched). Cysteine carbamidomethylation (+57.0215) and lysine TMT labeling (+229.1629) were specified as fixed modifications, and methionine oxidation (+15.9949), N-terminal protein acetylation (+42.0106), and TMT labeling of the peptide N terminus. For the whole proteome database search, TMT labeling on Serine residues was also specified as a variable modification. The search was restricted to fully tryptic peptides, allowing up to two missed cleavage sites. Phosphopeptide-enriched searches also included the phosphorylation modification of serine, threonine, and tyrosine residues (+79.9663). The search results were then processed using the Philosopher toolkit version v1.2.3 (da Veiga Leprevost et al., 2020), including PeptideProphet (Keller et al., 2002), PTMProphet (Shteynberg et al., 2019), and ProteinProphet (Nesvizhskii et al., 2003). The data was filtered to 1% PSM-level (for each 11-plex), and 1% protein-level (global) FDR using Philosopher filter command. TMT-Integrator version v1.0.4 (http://tmt-integrator.nesvilab.org/) was used for generation of quantification matrices as described previously (Clark et al., 2019; Djomehri et al., 2020), except its parameters were adjusted to process 11 channels, a minimum peptide probability of 0.5 for quantification, and minimum site localization probability of 0.75 (phosphopeptide-enriched datasets only). Quantification results (log2 ratios) were summarized at protein and gene levels, and for phosphopeptide enriched data also at the site-level.

Preprocessing of TMT proteomic data

8802 unique genes and 18235 phosphosites were identified and quantified from the global proteomic and phopshoproteomic experiments. Global normalization was performed on the gene-level abundance matrix (log2 ratio) for global proteomic and on the site-level abundance matrix (log2 ratio) for phosphoproteomic data. Specifically, each sample was shifted to have the same median, and scaled to have the same median absolute deviation.

We then applied an 'Intra TMT-multiplex t test' to detect and remove outlier TMT multiplexes for each protein/phosphosite. For each TMT multiplex, we performed t test of the protein/phosphosite abundance of samples inside against the protein/phosphosite abundance of samples outside the multiplex. TMT multiplexes with a p value lower than 10e-7 were flagged as outliers and removed from the dataset. Accordingly, a total of 164 and 156 multiplexes or 1612 and 1557 data points were removed in the global protein abundance and phosphosite datasets respectively.

Before performing any downstream analysis, we applied diagnosis-specific batch correction on both global and phospho abundance to remove the technical difference between different TMT 10-plex. For each data type, batch correction was performed on the subset of markers with more than 50% observed in at least one of the subtypes. After filtering markers with missing rates > 50% in all subtypes, there were 6429 genes and 4988 phosphosites with an overall missing rates of 20.2% and 40.8% for the diagnosis-wide global protein and phospho abundance datasets respectively.

As a first step to batch correction, we performed KNN imputation separately on the data from each diagnosis using the ''impute.knn'' function from the ''impute'' R package. After merging the data across diagnoses, we then applied the R tool ComBat, with the tumor diagnosis as covariate to remove batch effects (Johnson et al., 2007). Finally, we replaced the missing data structure from before KNN imputation.

For the formal imputation of missing values, we adopted a novel tool DreamAI (Ma et al., 2020; https://github.com/WangLab-MSSM/DreamAI), an ensemble algorithm developed during the NCI-CPTAC Dream Proteomics Imputation Challenge (https://www.synapse.org/#!Synapse:syn8228304/wiki/413428). Imputation was done: 1) separately on the data from each tumor type, and 2) across the entire dataset including all tumors. Tumor-subtype specific imputation was done for the subset of markers with missing < 50% in each subtype. Subtype-wide imputation was done for the subset of markers that appeared in at least 50% of samples in any one subtype (the same set of markers used in the batch correction). Finally, for the phospho abundance dataset, we filtered out 440 additional markers associated with cold-regulated ischemia genes.

QC check for proteogenomic profiles

Integration of these multi-layers of omics data enhances our understanding about complex molecular mechanisms in biological systems. However, unintended errors in annotations and sample sables often occur in generation or management of large-scale data (Alyass et al., 2015). Since integrative analysis based on error-containing data could provide wrong scientific conclusions, data quality and sample-labeling check is a critical QC step before actual integration. In this study, we performed systematic quality control procedure to confirm that all annotations in clinical information and sample names are consistent as annotated.

1. *Diagnosis type check and filtering* Among the 226 samples, 7 samples were identified to have either incorrect or ambiguous histologic diagnosis based on an independent clinical report review and were removed from the downstream analysis.
2. *Gender label check* Expression of two gender representative genes, XIST and RPS4Y1 from chromosomes X and Y, respectively (Staedtler et al., 2013), were used to infer genders based on RNaseq data.
3. *Genotype mapping check based on WGS and RNaseq data* To ensure the highest data quality, genotype mapping analysis was performed to flag samples with potential contamination, low sequencing quality, or sample labeling issues. Specifically, using NGSCheckMate (Lee et al., 2017), the genotype correlation were compared between paired tumor WGS versus normal WGS as well as tumor WGS versus tumor RNaseq profiles for all patients. The tool utilized 20,000+ common SNP sites based on dbSNP138, and a stringent cutoff of 0.8 was applied to flag low-quality or contaminated samples.

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

4. *Proteo-genomic sample labeling mapping* We employed similar procedures applied to our recent kidney cancer study (Clark et al., 2019) to confirm that RNaseq, global proteomics, and phosphoproteomic data with the same labels were from the same individuals (Yoo et al., 2014). *Cis* pairs among global proteomics, phosphosite proteomic, and RNaseq data were determined based on their correlation strength (*cis* correlation > 0.6) and 832, 1341, and 521 pairs were selected for global-phosphosite, RNaseq-global, and RNaseq-phosphosite alignments, respectively. Then the values of the selected features (genes or proteins) were rank-transformed to evaluate sample-wise similarity scores. If two profiles of a sample are matched, the similarity score between the two profiles is expected to be significantly higher than the score of random pairs. Based on this approach, we identified potentially mis-aligned samples from pairwise alignment among global proteomic, phosphoproteomic, and RNaseq data.

For flagged samples, same subject tumor tissue and blood specimen DNA was further extracted and sent to Guardian Forensic Sciences (Abington, PA) for short tandem repeat (STR) testing using the GenePrint 24 assay (Promega, #B1870). Pattern of amplified polymorphic loci was used for matching analysis between tissue and blood for each case.

After filtering data files according to all above quality assessments, the resulting datasets consisting of 218 global proteomics profiles, 217 phosphoproteomics profiles, 188 RNaseq profiles, 200 mutation profiles and 190 CNV profiles were considered for downstream analyses (Table S1).

Proteomics experiment of the validation cohort

Protein Extraction and Lys-C/Trypsin Tandem Digestion

Approximately 50 mg of each of brain tumor tissues were cryopulverized and lysed separately in 800 μL of lysis buffer (6 M urea, 25 mM Tris, pH 8.0, 1 mM EDTA, 1 mM EGTA, 1:100 v/v Sigma protease inhibitor, 1:100 v/v Sigma phosphatase inhibitor cocktail 2, and 1:100 v/v Sigma phosphatase inhibitor cocktail 3). Lysates were precleared by centrifugation at 20,000 g for 10 min at 4°C and protein concentrations were determined by BCA assay and adjusted to approximately 1.5 μg/μL with lysis buffer. Proteins were reduced with 5 mM dithiothreitol for 1 h at 37°C, and subsequently alkylated with 10 mM iodoacetamide for 45 min at 25°C in the dark. Samples were diluted to 2 M urea concentration with 25 mM Tris, pH 8.0 and digested with Lys-C (Wako) at 1:50 enzyme-to-substrate ratio. After 2 h of digestion at 25°C, aliquot of sequencing grade modified trypsin (Promega, V5117) at 1:25 enzyme-to-substrate ratio was added to the samples and further incubated at 25°C for 14 h. The digested samples were then acidified with 100% formic acid to 1% of final concentration of formic acid and centrifuged for 15 min at 1,500 g to clear digest from precipitation. Tryptic peptides were desalted on C18 SPE (Waters tC18 SepPak, WAT054925) and dried using Speed-Vac.

TMT-11 Labeling of Peptides

Desalted peptides from each sample were labeled with 11-plex TMT reagents. Peptides (400 μg) from each of the samples were dissolved in 80 μL of 50 mM HEPES, pH 8.5 solution, and mixed with 400 μg of TMT reagent that was dissolved freshly in 20 μL of anhydrous acetonitrile according to the optimized TMT labeling protocol described previously (Zecha et al., 2019). Channel 126 was used for labeling the internal reference sample (pooled from 100 adult GBM tumor and 10 GTEx normal samples (Wang et al., 2020)) throughout the sample analysis. After 1 h incubation at RT, 60 μL 50 mM HEPES pH8.5, 20% ACN solution was added to dilute the samples, and 12 μL of 5% hydroxylamine was added and incubated for 15 min at RT to quench the labeling reaction. Peptides labeled by different TMT reagents were then mixed, dried using Speed-Vac, reconstituted with 3% acetonitrile, 0.1% formic acid and desalted on tC18 SepPak SPE columns.

Peptide Fractionation by bRPLC

Approximately 3.5 mg of 11-plex TMT labeled sample was separated on a reversed phase Agilent Zorbax 300 Extend-C18 column (250 mm × 4.6 mm column containing 3.5-μm particles) using the Agilent 1200 HPLC System. Solvent A was 4.5 mM ammonium formate, pH 10, 2% acetonitrile and solvent B was 4.5 mM ammonium formate, pH 10, 90% acetonitrile. The flow rate was 1 mL/min and the injection volume was 900 μL. The LC gradient started with a linear increase of solvent B to 16% in 6 min, then linearly increased to 40% B in 60 min, 4 min to 44% B, 5 min to 60% B and another 14 of 60% solvent B. A total of 96 fractions were collected into a 96 well plate throughout the LC gradient. These fractions were concatenated into 24 fractions by combining 4 fractions that are 24 fractions apart (i.e., combining fractions #1, #25, #49, and #73; #2, #26, #50, and #74; and so on). For proteome analysis, 5% of each concatenated fraction was dried down and re-suspended in 2% acetonitrile, 0.1% formic acid to a peptide concentration of 0.1 mg/mL for LC-MS/MS analysis. The rest of the fractions (95%) were further concatenated into 12 fractions (i.e., by combining fractions #1 and #13; #3 and #15; and so on), dried down, and subjected to immobilized metal affinity chromatography (IMAC) for phosphopeptide enrichment.

Phosphopeptide Enrichment Using IMAC

Fe3+-NTA-agarose beads were freshly prepared using the Ni-NTA Superflow agarose beads (QIAGEN, #30410) for phosphopeptide enrichment. For each of the 12 fractions, peptides were reconstituted to 0.5 μg/μL IMAC binding/wash buffer (80% acetonitrile, 0.1% trifluoroacetic acid) and incubated with 10 μL of the bead suspension for 30 min at RT. After incubation, the beads were sequentially washed with 50 μL of wash buffer (1X), 50 μL of 50% acetonitrile, 0.1% trifluoroacetic acid (1X), 50 μL of wash buffer (1X), and 50 μL of 1% formic acid (1X) on the stage tip packed with 2 discs of Empore C18 material (Empore Octadecyl C18, 47 mm; Supleco, 66883-U). Phosphopeptides were eluted from the beads on C18 using 70 μL of elution buffer (500 mM K2HPO4, pH 7.0). Sixty microliter of 50% acetonitrile, 0.1% formic acid was used for elution of phosphopeptides from the C18 stage tips after two washes with 100 μL of 1%

formic acid. Samples were dried using Speed-Vac and later reconstituted with 12 μL of 3% acetonitrile, 0.1% formic acid for LC-MS/MS analysis.

LC-MS/MS Analysis

Fractionated samples prepared for whole proteome and phosphoproteome analysis were separated using a nanoACQUITY UPLC system (Waters) by reversed-phase HPLC. The analytical column was manufactured in-house using ReproSil-Pur 120 C18-AQ 1.9 μm stationary phase (Dr. Maisch GmbH) and slurry packed into a 25-cm length of 360 μm o.d. x 75 μm i.d. fused silica picofrit capillary tubing (New Objective). The analytical column was heated to 50°C using an AgileSLEEVE column heater (Analytical Sales and Services). The analytical column was equilibrated to 98% Mobile Phase A (MP A, 0.1% formic acid/3% acetonitrile) and 2% Mobile Phase B (MP B, 0.1% formic acid/90% acetonitrile) and maintained at a constant column flow of 200 nL/min. The sample was injected into a 5 μL loop placed in-line with the analytical column which initiated the gradient profile (min:%MP B): 0:2, 1:6, 85:30, 94:60, 95:90, 100:90, 101:50, 110:50. The column was allowed to equilibrate at start conditions for 30 minutes between analytical runs.

MS analysis was performed using an Orbitrap Fusion Lumos mass spectrometer (ThermoFisher Scientific). The whole proteome and phosphoproteome samples were analyzed under identical conditions. Electrospray voltage (1.8 kV) was applied at a carbon composite union (Valco Instruments) coupling a 360 μm o.d. x 20 μm i.d. fused silica extension from the LC gradient pump to the analytical column and the ion transfer tube was set at 250°C. Following a 25-min delay from the time of sample injection, Orbitrap precursor spectra (AGC 4 × 105) were collected from 350-1800 m/z for 110 min at a resolution of 60K along with data dependent Orbitrap HCD MS/MS spectra (centroid) at a resolution of 50K (AGC 1 × 105) and max ion time of 105 ms for a total duty cycle of 2 s. Masses selected for MS/MS were isolated (quadrupole) at a width of 0.7 m/z and fragmented using a collision energy of 30%. Peptide mode was selected for monoisotopic precursor scan and charge state screening was enabled to reject unassigned 1+, 7+, 8+, and > 8+ ions with a dynamic exclusion time of 45 s to discriminate against previously analyzed ions between ± 10 ppm.

Quantification of TMT Whole Proteomic Data

The Thermo RAW files were processed with mzRefinery to characterize and correct for any instrument calibration errors, and then with MS-GF+ v9881 (Gibbons et al., 2015; Kim et al., 2008; Kim and Pevzner, 2014) to match against the RefSeq human protein sequence database downloaded on June 29, 2018 (hg38; 41,734 proteins), combined with 264 contaminants (e.g., trypsin, keratin). The partially tryptic search used a ± 10 ppm parent ion tolerance, allowed for isotopic error in precursor ion selection, and searched a decoy database composed of the forward and reversed protein sequences. MS-GF+ considered static carbamidomethylation (+57.0215 Da) on Cys residues and TMT modification (+229.1629 Da) on the peptide N terminus and Lys residues, and dynamic oxidation (+15.9949 Da) on Met residues for searching the global proteome data (Monroe et al., 2008). Next, PSMs passing the confidence thresholds described above were linked to the extracted reporter ion intensities by scan number. The reporter ion intensities from different scans and different bRPLC fractions corresponding to the same gene were grouped. Relative protein abundance was calculated as the ratio of sample abundance to reference abundance using the summed reporter ion intensities from peptides that could be uniquely mapped to a gene. The pooled reference sample was labeled with TMT 126 reagent, allowing comparison of relative protein abundances across different TMT-11 plexes. The relative abundances were log2 transformed and zero-centered for each gene to obtain final relative abundance values. Small differences in laboratory conditions and sample handling can result in systematic, sample-specific bias in the quantification of protein levels. In order to mitigate these effects, we computed the median, log2 relative protein abundance for each sample and re-centered to achieve a common median of 0.

Quantification of Phosphopeptides

Phosphopeptide identification for the phosphoproteomic data files were performed as in the whole proteome data analysis described above (e.g., peptide level FDR < 1%), with an additional dynamic phosphorylation (+79.9663 Da) on Ser, Thr, or Tyr residues. The phosphoproteome data were further processed by the Ascore algorithm (Beausoleil et al., 2006) for phosphorylation site localization, and the top-scoring sequences were reported. For phosphoproteomic datasets, the TMT-11 quantitative data were not summarized by protein but left at the phosphopeptide level. To account for sample-specific biases in the phosphoproteome analysis, we applied the correction factors derived from median-centering the whole proteomic dataset. Preprocessing of the proteomic tables of the Project Hope sample analysis were performed in the same fashion in the pediatric sample analysis described above.

Targeted Mass Spectrometry Methods

For targeted mass spectrometry measurements, tissue lysates were reduced, alkylated with iodoacetamide, and digested by the addition of trypsin at a 1:50 trypsin:protein ratio (by mass), as previously described (Whiteaker et al., 2018). After 2 hours, a second trypsin aliquot was added at a 1:100 trypsin:protein ratio and incubated overnight at 37°C with shaking. After 16 hours, the reaction was quenched with formic acid (final concentration 1% by volume). A mix of stable isotope-labeled peptide standards was added to the digest at 80 fmol/mg per peptide.

Peptide immunoaffinity enrichment was performed as previously described (Zhao et al., 2011), using a mixture of 50 antibodies crosslinked on protein G beads targeting 75 peptides (21 modifications, 40 proteins). LC-MRM was performed as previously described (Whiteaker et al., 2018).

Targeted MRM Assay Characterization

Response curves were generated in a background matrix of pooled brain tumor lysates. Five hundred microgram aliquots of the pooled lysate were digested by trypsin, and the heavy stable isotope-labeled peptides were added to aliquots in triplicate by serial

dilution covering the amounts 1000, 200, 40, 8, 3.2, 1.28, 0.512, 0.205 fmol/mg with light spiked into the pool at 80 fmol/mg. Blanks were prepared using a background matrix with light peptide (no heavy spike). All points were analyzed in triplicate (including peptide addition, immunoaffinity enrichment, and mass spectrometry). Data analysis was performed using Skyline. The Lower Limit of Quantification (LLOQ) was obtained by empirically finding the lowest point on the curve with a CV < 20% in the curve replicates. All measurements were filtered by the LLOQ (i.e., all measurements were required to be above the LLOQ). The upper limit of quantification (ULOQ) was determined by the highest concentration point of the response curve that was maintained in the linear range. For curves that maintained linearity at the highest concentration measured, the ULOQ is a minimum estimate.

Repeatability was determined using the same pooled lysate matrix used to generate the response curves with heavy peptides spiked in at three concentrations (0.8, 80, 800 fmol/mg) and light peptides added at 200 fmol/mg. Complete process triplicates (including digestion, capture, and mass spectrometry) were prepared and analyzed on five independent days. Intra-assay variation was calculated as the mean CV obtained within each day. Inter-assay variation was the CV calculated from the mean values of the five days.

Targeted MRM Data Results

The median LLOQ was 1.6 fmol/mg and the median linear range > 2.8 orders of magnitude. In repeatability experiments, the median CV at the medium spike level was 8.6% (intra-assay CV) and 26% (inter-assay CV).

Each data point is the peak area ratio (light:heavy) filtered by the LLOQ. The unfiltered data points are also available in the table. For each sample 500 ug aliquots were analyzed in complete process replicate (including digestion, capture, and mass spectrometry). The number of replicates available for processing was determined by the amount of lysate available. Overall, 68 out of the 75 peptide analytes were detected in > 50% of the samples above the LLOQ. Five peptides were not detected in any samples (TNF10.pan.NGELVIHEK, ATM pS2996, ATMpS367, RIF1 pS1542, RIF1.pan.ASQGLLSSIENSESDSSEAK). For peptides with replicates available, the median CV was 12.9%. The correlation of peak area ratios for peptides originating from the same protein was high (RPTOR: R2 = 0.9047, ERBB2: R2 = 0.9536, K25: R2 = 0.9964) indicating good quality for multiple measurements of the same protein.

QUANTIFICATION AND STATISTICAL ANALYSIS

Kinase Activity Score Calculation

Substrates of every kinase were collected from the PhosphoSitePlus database (version 052819). We only considered kinases with at least five substrates observed in our phosphoproteomic data. To calculate the kinase activity score for each sample, we run a Wilcoxon rank sum test comparing the abundance of substrates of a particular kinase with that of the remaining phosphosites observed in our data. This test was performed for each kinase and each of the 209 samples (i.e., excluding the post-mortem samples). The normalized test statistic of the Wilcoxon test was utilized as the activity score for each kinase.

Consensus clustering analysis

Proteomic cluster

Consensus clustering was performed to identify proteo-typical clusters of childhood brain tumors. Based on gene level global proteomics data, features (genes) were first filtered according to the coefficient of variation (CV) and standard deviation (SD) across samples. Specifically, the CV was calculated using the raw intensity data; features with CV less than 0.1 were filtered resulting in the exclusion of 583 genes for consideration; finally, 3000 genes with the highest standard deviation across 218 samples were selected for clustering.

Consensus clustering was performed using the ConsensusClusterPlus package in R (Wilkerson and Hayes, 2010). Prior to clustering the data matrix was scaled so that each peptide had a mean 0 and a sd of 1 across samples. K-means clustering based on an Euclidean distance metric was conducted across 500 repetitions for cluster numbers ranging from 2 through 10 using otherwise default parameters.

Phosphoproteomic and transcriptomic cluster

To compare proteomic clusters with those derived from alternative -omic data types, RNA-seq and phosphosite clusters were identified using a similar procedure. Specifically, phosphosite data from 217 samples were clustered using the 3000 phosphosites with the largest SD after first filtering 830 phospho-sites with CV > 0.1. For the clustering of RNA-seq data, ½ of the genes with the highest standard deviation were selected corresponding to 9104 features from 188 tumor samples.

Inspection of the CDF distribution, as well as patterns of concordance across data types and with histological diagnosis, led to the selection of 8 clusters for further analysis.

Comparison across single-omic clusters

To evaluate and compare the cohesiveness of allocations derived using single-omic clustering, we utilized silhouette scores. Silhouette scores, which measure the similarity of a given sample to the other samples in the same cluster, were calculated using the silhouette function from the 'cluster' package in R. For each single-omic dataset, the Euclidean distance matrix used in the consensus clustering and the respective single-omic clustering allocations were included as inputs to the silhouette function.To further compare clustering allocations across single-omic datasets and with histological diagnoses, the percentage of each diagnostic type falling into each cluster was calculated for each set of allocations. This data is summarized in Figure S1B.

Survival analysis for proteomic clusters

The association of proteomic clusters and overall survival was evaluated using a Cox model based on 198 patients with surgical samples and overall survival information (Figure 1C). Overall survival values were truncated to a maximum of 3750 days.

Stemness Score

The stemness indices are used for assessing the degree of oncogenic dedifferentiation, as previously described (Malta et al., 2018). Stemness can be considered to be the ability of the tumor to phenocopy a normal stem cell. Higher values for stemness indices were associated with biological processes active in cancer stem cells and with greater tumor dedifferentiation, as reflected in histopathological grade (Malta et al., 2018). Recently several signaling pathways associated with stemness have been reported for each of the mentioned CBTN PBT diagnoses (Chang et al., 2017; Liu et al., 2020; Meel et al., 2018).

Stemness scores were calculated as previously described (Malta et al., 2018). First, we used MoonlightR (Colaprico et al., 2020) to query, download, and preprocess the pluripotent stem cell samples (ESC and iPSC) from the Progenitor Cell Biology Consortium (PCBC) dataset (Daily et al., 2017; Salomonis et al., 2016). Second, to calculate the stemness scores based on mRNA expression, we built a predictive model using one-class logistic regression (OCLR) (Sokolov et al., 2016) on PCBC dataset. To calculate mRNA based stemness index (mRNASi), we used the FPKM (Fragments Per Kilobase Million) mRNA expression values for all the 188 CBTN PB tumors. We used the function TCGAanalyze_Stemness from the package TCGAbiolinks (Colaprico et al., 2016) and following our previously-described workflow (Mounir et al., 2019), with "stemSig" argument set to PCBC_stemSig.

Proliferative Index

Proliferative index was calculated based on gene expression data of 40 genes contained in the proliferation gene signature from Yuan et al., (Yuan et al., 2018). The proliferative index was computed via sample-specific gene set enrichment analysis (ssGSEA) score using the package GSVA (Hänzelmann et al., 2013).

Investigation of two subtypes of Cranio

To better characterize the biological features differentiating the CP allocated to different proteomic clusters and to investigate the hypothesis that CP allocated to the *Cranio/LGG BRAF^{V600E} (C4)* may respond to MEK inhibitor treatment, regression analyses were performed using the global proteomic data to identify markers differentially expressed between C4 and C8 with CTNNB1 status accounted for as a covariate. Gene set enrichment tests were performed for a MEK inhibition response signature based on 15 genes overlapping between our global proteomic data and a 52 member geneset previously reported to be perturbed by MEK inhibitor treatment in multiple cancer cell lines with *BRAF^{V600E}* (Pratilas et al., 2009). The MEK inhibition response gene set was found to be significantly enriched of proteins upregulated in C4 (pvalue = 0.05), as is illustrated in the volcano plot in Figure S1C.

Proteomic Cluster Signature

To identify proteins and phospho-site markers associated with proteomic clusters, a multiple regression was performed using protein/phospho-site abundances as responses, and binary indicators representing the 8 proteomic-clusters as regressors, with age of specimen diagnosis, gender, as well as treatment and clinical status at sample collection included as covariates. Model fitting was performed without an intercept so that the resulting betas are interpretable as a mean shift relative to all tumors. Results from cluster-specific association testing performed on 6429 protein (N = 218), 4548 phosphosite (N = 218), and 18209 gene expression (N = 188) features are reported in Table S1.

Pathway analysis for proteomic clusters

To better characterize proteomic clusters, we sought to identify the biological pathways distinctly associated with each. First, using the R package ConsensusClusterPlus (Wilkerson and Hayes, 2010), genes were clustered based on a Z-score matrix (8 columns) summarizing the cluster-specific regression analysis results based on proteomics data. Specifically, each row of the Z-score matrix represents the estimated mean shift of a given protein's abundance in each of the 8 proteomic clusters. Considering between 10-20 clusters, K-means clustering was performed to group genes (N = 6353) whose protein abundances were significantly associated (FDR < 0.05) with at least one proteomic cluster. Pathway enrichment was performed to test for overrepresentation of biological pathway/gene-set members in each gene group using a one-tailed Fisher's Exact Test. The final number of gene groups (k = 14) was chosen to maximize the number of significant pathway associations based on the Hallmark gene sets from MSigDB (Liberzon et al., 2011, 2015) downloaded from https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp on 02/14/2019. Based on the 14 groups of gene so selected, a comprehensive pathway analysis was further performed using GO (Ashburner et al., 2000), Biocarta, KEGG (Kanehisa et al., 2017), Hallmark (Liberzon et al., 2015), and Reactome (Fabregat et al., 2018) gene set collections.

Pathway Consolidation via Sumer

Gene set enrichment results can be difficult to interpret due to significant redundancy of gene membership across collections of gene sets. To aid in the interpretation and reduce redundancy of pathway results, we utilized the Sumer tool (Savage et al., 2019). This tool uses an affinity propagation algorithm to cluster similar pathway gene sets into largely distinct modules. Sumer was run using

-log10 p values from the Fisher test of gene cluster enrichment as weights. Consolidated pathway modules for each gene cluster were identified, based on the top 50 pathways by weight (Table S1); a subset of these along with other pathways with biological relevance in cancer were selected for display in Figure 1D.

Immune subtype identification

The abundance of 64 different cell types were computed via xCell based on transcriptomic profiles (Aran et al., 2017). Therefore, for this analysis, 182 pediatric brain tumor samples with mRNA data were utilized excluding post-mortem samples. Table S2 contains the final score computed by xCell of different cell types. Consensus clustering was performed based on only cells which were detected in at least 5% of the patients (adjusted p value < 1%). This filtering resulted in 35 cell types. A Microglia signature was derived as ssGSEA score (Hänzelmann et al., 2013) based on the following microglia-specific markers: *P2RY12, TMEM119, SLC2A5, TGFBR1, GPR34, SALL1, GAS6, MERTK, C1QA, PROS1, CD68, ADGRE1, AIF1, CX3CR1, TREM2 and ITGAM* (Butovsky et al., 2014; Crotti and Ransohoff, 2016; Haage et al., 2019; Solga et al., 2015). Based on these 36 signatures, consensus clustering was performed in order to identify groups of samples with similar immune/stromal characteristics. Consensus clustering was performed using the R packages ConsensusClusterPlus (Wilkerson and Hayes, 2010) based on z-score normalized signatures. Specifically, 80% of the original pediatric brain tumor samples were randomly subsampled without replacement and partitioned into 5 major clusters using the Partitioning Around Medoids (PAM) algorithm, which was repeated 200 times (Wilkerson and Hayes, 2010; Figure 2A; Table S2).

single cell RNaseq deconvolution Analysis

We have applied the tool Music (Wang et al., 2019) trained on single cell sequencing data from Darmanis et al. (2017) to all the mRNA expression values for all 182 tumors considered for the immune subtype analysis. We used the function TCGAanalyze_scRNA (tool = Music, data = GSE84465) from the package TCGAbiolinks (Colaprico et al., 2016) to query, download and prepare the data from Darmanis et al. (2017) and subsequently obtain microglia, neuronal and oligodendrocytes cell type composition in these tumors. scRNA data were normalized following previously-described workflow (Lun et al., 2016).

Tumor Purity, Stromal and Immune Scores

Besides xCell, we utilized ESTIMATE (Yoshihara et al., 2013) to infer immune and stromal scores based on gene expression data (Table S2). To infer tumor purity, TSNet was utilized (Petralia et al., 2018; Table S2).

Differentially Expressed Genes and Pathway

Genes upregulated in each of the five immune clusters were identified based on gene expression data, global proteomic and phosphoproteomic data. For this analysis, imputed proteomic and phosphoproteomic data were utilized. For each data type, every feature vector was normalized by subtracting the mean and dividing by the standard deviation across 182 samples. Then, for each data type, the expression level of gene/protein/phosphosite j was modeled via

$$x_{i,j} = \sum_{k=1}^{5} \beta_{k,j} 1(i \in I_k) + \varepsilon_{i,j} \qquad \text{(Equation 1)}$$

with $\varepsilon_{i,j} \sim N(0,\sigma_j)$, I_k being the set of samples belonging to the k-th immune cluster, $1(A)$ being an indicator function equal to 1 if the event A occurs and 0 otherwise, $\beta_{k,j}$ being the coefficient capturing the association between gene j and the k-th immune group. Benjamini adjusted p values (Benjamini and Hochberg, 1995) can be found in Table S2. For each immune cluster, considering the set of genes upregulated with Benjamini's adjusted p value lower than 1%, a fisher exact test was implemented to derive enriched pathways. For this analysis, pathways from the Reactome (Fabregat et al., 2018), KEGG (Kanehisa et al., 2017), Hallmark (Liberzon et al., 2015) and GO (Ashburner et al., 2000) databases were considered and as background the full list of gene/proteins observed under each data type was utilized. For phosphorylation data, a gene was considered upregulated if at least one substrate of the gene was found upregulated based on phosphorylation data at 1% FDR. The pathway analysis results for different data types are contained in Table S2B. Figure 2B contains key pathways significant at 10% FDR for different data types. Given their similarity in terms of enriched pathways, the two cold immune clusters (i.e., Cold-medullo and Cold-mixed) were combined into one category and pathways upregulated in both clusters at 10% FDR were reported in Figure 2B. Pathway scores for 182 pediatric brain tumor samples were computed based on ssGSEA using the R package GSVA and included in Figure 2A (Hänzelmann et al., 2013).

Microglia and Macrophage Polarization in LGG

Microglia polarization signatures were constructed with ssGSEA (Hänzelmann et al., 2013) using RNaseq measurements based on genes described in recent literature (Dello Russo et al., 2017; Fumagalli et al., 2018; Krasemann et al., 2017). Specifically, the following gene sets were considered: Proinflammatory (M1) = (*IL1B, TLR4, TNF, NOS2, APOE, CLEC7A, LGALS3, GPNMB, ITGAX, SPP1, CCL2, FABP5, CYBB*); Anti-inflammatory (M2) = (*COQ7, IL4, IL13, IL10, ARG1, TGFB1, SMAD3, HEXB, P2RY12, MERTK, ENTPD1, TMEM119, TGFBR1, CD163, CD206*). M2-0.65*M1 difference was used for Figure 2G.

Immune association with BRAF Status in LGG

For this analysis, we consider the 35 immune/stromal signature from xCell and the microglia signature utilized to perform the consensus clustering, microglia M1 and M2 signatures and antigen presenting machinery Class I and Class II signature. Antigen presenting machinery signature Class I was derived via ssGSEA score (Hänzelmann et al., 2013) using gene expression measurements of *HLA-A*, *HLA-B* and *HLA-C* genes; while class II signature based on the gene expression of *HLA-DPA1*, *HLA-DPB1*, *HLA-DRA*, *HLA-DRB1*, *HLA-DRB5* and *HLA-DQB1*. Each signature was normalized to z-score and then was modeled as function of *BRAF* status (i.e., BRAF wild-type, BRAF fusion and BRAF v600E) via linear model. Table S2 reports p values for association passing 10% FDR.

iProFun Based Cis Association Analysis

We investigated the functional molecular quantitative traits (mRNA, protein, and phosphoprotein abundances) perturbed by CNV, using an integrative analysis tool iProFun (Song et al., 2019). iProFun jointly models the multi-omics outcomes, and enjoys largely enhanced power for detecting significant cis-associations shared across different omics data types; and it also achieved better accuracy in inferring cis-associations unique to certain type(s) of molecular trait(s). Specifically, we considered three functional molecular quantitative traits (mRNA expression levels, global protein abundances, and phosphopeptide abundances) for their associations with CNV measured by log ratios. After removing post-mortem samples, we collected 168 pediatric brain tumor samples with all four platforms measured, and performed iProFun on these samples. Samples from different biopsies of the same subject (e.g., from initial tumor and progressed tumor) were both considered in the analysis. The mRNA expression levels were available for 18,209 genes, the global protein abundance measurements were available for 6,429 genes, the phosphopeptide abundance was available for 4,518 peptides from 1,958 genes, and the CNVs were obtained for 19,374 genes, respectively. All data types were preprocessed to eliminate potential issues for analysis such as batch effects, missing data and major unmeasured confounding effects. For this analysis, imputed proteomic and phosphoproteomic tables were utilized. The mRNA expression levels, global protein and phosphoprotein abundances were also normalized to standard normal distribution. To account for potential confounding factors, we considered age, gender, tumor purity, tumor diagnosis, treatment status at collection and somatic mutation. Tumor purity was determined using TSNet from RNA-seq data as described above.

The iProFun procedure was first applied to a total of 1622 genes measured across all 4 data types (mRNA, global protein, phosphoprotein, CNV). Specifically, we started with traditional linear regression for each of the three outcomes separately: mRNA ~CNV + covariates, global ~CNV + covariates, and phospho ~CNV + covariates. Then, the association summary statistics from regressions was taken as input for iProFun to call posterior probabilities of belonging to each of the eight possible configurations (“None,” “mRNA only” “global only,” “phospho only” “mRNA & global,” “mRNA & phospho,” “global & phospho” and “all three”) and to determine significance associations.

Table S3 presents the significant genes that pass the following three criteria: (1) the satisfaction of biological filtering procedure, (2) posterior probabilities > 75%, and (3) empirical false discovery rate (eFDR) < 10%. Specifically, the biological filtering criterion requires that CNV presents positive associations with all the types of molecular QTs. Second, a significance was called only if the posterior probabilities > 75% of a predictor being associated with a molecular QT, by summing over all configurations that are consistent with the association of interest. For example, the posterior probability of a CNV being associated with mRNA expression levels was obtained by summing up the posterior probabilities in the following four association patterns – “mRNA only,” “mRNA & global,” “mRNA & phospho” and “all three,” all of which were consistent with CNV being associated with mRNA expression. Lastly, we calculated empirical FDR via 100 permutations per molecular QTs by shuffling the label of the molecular QTs, and requested empirical FDR (eFDR) < 10% by selecting a minimal cutoff value of alpha that 75% < alpha < 100%. The eFDR is calculated by:

> eFDR = (Averaged No. of genes with posterior probabilities > alpha in permuted data) / (Averaged No. of genes with posterior probabilities > alpha in original data).

In total, we identified 515 genes whose CNV showed cascading cis-regulation of their mRNA expression levels, global protein and phosphopeptide abundances.

Similarly, iProFun was applied to a total of 6183 genes measured across all 3 data types (mRNA, global protein, CNV) for their *cis* regulatory patterns in tumors, and 1541 genes whose CNV showed cascading cis-regulation of their mRNA expression levels and global protein abundances. To further visualize the cascading genes from iProFun analysis, we selected a subset of cascading genes which have adequate copy number activity in any of the diagnosis subtypes, and marginally differentiated in protein/phosphosite abundance across different copy number status in the same subtype. We define copy number activity by comparing CNVs with the standard deviation across all samples on the same location: cnv over 1-fold SD was regarded as gain and below negative 1-fold SD regarded as loss. Adequate copy number activity was defined with the total proportion of gain and loss over 25% and either category including at least 2 samples. After categorizing CNV with 3 groups: gain/normal/loss, we tested if protein/phospho abundance was differentially distributed with contrast on gain-to-normal or loss-to-normal by two sample Wilcoxon-test. Genes with p value below 0.1 in the test under one of the contrasts were indicated as marginal associated with CNV. All of the selected cascade genes were labeled along the genome in Figure S3D and those genes also being reported as druggable targets or oncogenes were listed with their symbols on the same plot.

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

Cis-regulation of Somatic Mutations

We considered genes whose mutation rate is prevalent (at least 6 mutations across 200 tumors) to investigate their associations with their *cis* mRNA, global and phosphoprotein abundances. A total of 46 genes were therefore considered in this association analysis. For each mutation, we considered the existence of any types of mutation (Yes/No) as primary predictor, mRNA/protein/phosphosite abundance as outcome, and CNV, age, gender, tumor purity and tumor diagnosis types as covariates, and performed linear regressions for their associations.

Trans association analysis

For each tumor diagnosis, we investigated the trans associations of its abundant genomic events on all measured mRNA, protein, phosphosite levels that pass QC procedures. For each subtype we calculated the chromosome arm-level cnv activity level using similar criterion as described in iProfun analysis. We compared arm-level CNVs in a specific subtype with their standard deviation across all samples on the same arm, calculated total proportion of gain (over 1-fold SD) and loss (below negative 1-fold SD). The proportion over 25% represented adequate activity in the chromosome arm region. We selected 5 diagnosis subtypes which contains at least 1 active CNV region to test trans associations. We also consider the trans association between two signature mutations and their highly enriched subtypes. Specifically, we considered association on chromosome arm *6q*, *17p*, *17q*, and *22q* in atypical teratoid rhabdoid tumor; *CTNNB1* mutation and *11p* in CP; *1q* and *8q* in EP; *7p* in ganglioglioma; *NF1* mutation, *1q, 6q, 7p, 9p, 9q, 11p, 13q, 14q, 16q, 17q* and *21p* in high grade astrocytoma; *1q, 7p, 7q, 8q, 10q, 11p, 11q, 16q, 17p, 17q* and *18p* in MB. For each of these genomic events, we investigated their association with all mRNA levels, protein abundances and phosphosite abundances among patients with the corresponding diagnosis, using unadjusted linear regression. Additional covariates were not considered due to small sample sizes in subtypes. We reported significant trans associations if FDR < 0.1.

To further understand the biological impact of the trans-regulations, we tested enrichment of positively regulated or negatively regulated gene set in pathways with fisher exact test. Enrichment test was performed on both RNA trans-regulated genes and protein trans-regulated genes. In this test, pathways from the Reactome (Fabregat et al., 2018), KEGG (Kanehisa et al., 2017) and GO (Ashburner et al., 2000) databases were considered and as background the full list of gene/proteins observed under each data type was utilized. Some pathways were enriched by trans-regulated genes in protein but not in RNA. For example, members of the "Cell Cell Contact Zone" pathway (purple) are enriched in the set of proteins upregulated in *CTNNB1* mutant samples; while "Coagulation" pathway is enriched in proteins downregulated in *CTNNB1* mutant samples.

Kinase Activity across different histologies

Kinase activity scores were calculated following the strategy illustrated in section "Kinase Activity Score Calculation." Table S4 contains the kinase activity for all diagnosis. For this analysis, we used proteomics and phosphoproteomic imputed data. The activity of each kinase was modeled as a function of the diagnosis indicator and the treatment information via a linear regression. Given the impact of post-mortem collection on proteogenomic data, post-mortem samples have been excluded from the analysis. P values were adjusted for multiple comparison via Benjamini & Hochberg adjustment (Table S4). In addition, for each diagnosis, the correlation between kinase activity and global abundance is reported in Table S4.

Kinase-Substrates Association

To discover the phosphorylation events that were relevant to pediatric brain tumors, we utilized the phosphosite-level data to examine the overall relationship between kinase global abundance and phospho-abundance with targeted sites. Given the impact of post-mortem collection on proteogenomic data, post-mortem samples have been excluded from the analysis. For this analysis, we used proteomic and phosphoproteomic imputed data. For each diagnosis, only kinases and phosphosites measured for at least 50% of the samples have been considered in the analysis. Since ATRT and MB tumors were merged into one group of samples, only kinases and phosphosites observed in at least one diagnosis (i.e., ATRT and MB) for more than 50% of the samples were utilized. For this analysis, experimentally validated kinase-substrate associations were considered from PhosphoSitePlus (Table S4; Hornbeck et al., 2015). This filtering resulted in a total number of 540 kinase-substrates possible associations between 82 unique kinases and 267 unique substrates (Table S4). Then, each phosphosite abundance was modeled as a function of targetable kinases via a multivariate linear regression adjusting for treatment information. When both phospho-abundance and global-abundance data were available for a particular kinase, the data type with higher correlation with the targeted site was considered in the model. For each diagnosed subtype, we adjusted for multiple comparisons via permutation technique. In particular, for each permutation, we run the multivariate analysis after randomly permuting the sample order of the abundance of the targeted site. Repeating this analysis for 200 permutations, we generated the distribution of p values under the null hypothesis of no-association and utilized this distribution to compute FDR (Tusher et al., 2001). Only associations passing an FDR adjustment of 10% were reported as significant (Table S4; Figure S4A). Note, given the small sample size of the ATRT and MB cohorts, their shared identity as embryonal tumors and their proteomic similarity (Figure 1D), ATRT and MB samples were combined to form the ATRT/MB group in this analysis.

Validation of kinase-phospho associations

Kinase-phospho associations detected in HGG were validated using proteomic and phosphoproteomic data for 23 additional high-grade glioma samples. This additional data is reported in Table S4.

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

BRAF mutation association analysis for LGG

A regression analysis was performed to compare the abundance of wild-type to each mutant type (fusion or point), with age of diagnosis and diagnosis type (initial or progressed) as covariates (Table S5). For this analysis, the LGG diagnosis-specific imputed tables (N = 93) were used, including of 5629 and 3437 markers for protein and phosphoproteomic data respectively. Association analysis was also performed for 85 LGG samples across 18209 transcripts.

Pathways associated with BRAF status in LGG

Wilcoxon enrichment analysis (WEA) was used to test for association between pathway genesets and *BRAF* status among LGG samples based on regression results from RNA and protein data. Gene set enrichment was conducted across multiple collections of genesets, including GO (Ashburner et al., 2000), Biocarta, KEGG (Kanehisa et al., 2017), Hallmark (Liberzon et al., 2015), and Reactome (Fabregat et al., 2018). These collections were downloaded from https://www.gsea-msigdb.org/gsea/msigdb/index.jsp on 2/14/2019. Gene sets with less than 5 or more than 250 member genes were excluded. A total of 4795 and 6215 genesets fitting this criterion were tested for enrichment in proteomic and RNA datasets respectively.

Consolidation pathways via Sumer

To help identify pathways distinctly associated with each mutation type and to consolidate redundant pathway results, Sumer software was utilized (Savage et al., 2019). The -log10 signed p value derived from a Z-test comparing mean ssGSEA scores (Hänzelmann et al., 2013) between mutant types was used as the pathway weight when running Sumer. Consolidated pathway modules are shown based on the top 150 pathways by weight (Table S5); a subset of these along with other pathways with biological relevance in cancer were selected for discussion in the main text.

Phosphoproteomic Co-expression Network in LGG

Network inference was utilized to characterize co-expression patterns among phosphorylation sites in LGG. The co-expression network was estimated based on phosphosite level data through a random-forest based algorithm (Petralia et al., 2016; Petralia et al., 2015). In particular, co-expression networks were estimated using LGG-specific imputed phosphorylation data. In order to deal with the fact that sites mapping to the same protein are usually correlated, we only modeled each site as function of sites mapping to other proteins. Let p be the total number of sites measured for n samples. Specifically, let x_{ij}^{s} be the abundance of the j-th site mapping to the s-th protein for the i-th sample. Then, x_{ij}^{s} was modeled as a function of other protein phosphosites, i.e., $\{x_{ij}^{k}\}_{k \neq s}$, via random forest. In order to derive the final unweighted networks, a proper cut-off value was chosen via permutation techniques (Fruchterman and Reingold, 1991; Petralia et al., 2016). Specifically, 50 permutations and an FDR cut-off of 1E-4 was considered to derive the final network (Table S5). For the visualization of network modules (Figures 5C and S5C) the software iCAVE (Kalayci and Gumus, 2018; Liluashvili et al., 2017) and Cytoscape (Shannon et al., 2003) were utilized. Force-directed layout algorithm (Fruchterman and Reingold, 1991) was applied to calculate initial positioning of nodes, node positions were then manually adjusted for visual concerns.

Network modules associated with BRAF status

Based on the network topology, network modules were identified using an algorithm based on edge betweenness score (Csárdi and Nepusz, 2006; Newman and Girvan, 2004). A total number of 18 network modules containing more than 20 phosphosites were derived (Table S5). Given a network module, the association with $BRAF^{V600E}$ and $BRAF^{Fusion}$ was found via fisher-exact test. In particular, a one-sided fisher exact test was performed to find modules enriched of sites differentially expressed between $BRAF^{Fusion}$ and *BRAF* wild-type and $BRAF^{V600E}$ and *BRAF* wild-type at 10% FDR (Table S5). P values were then adjusted for multiple comparison via Benjamini-Hochberg adjustment (Benjamini and Hochberg, 1995).

Network modules and druggable kinases

For this analysis, we considered kinases, which have been used, in clinical trials based on Open Targets database (https://www.targetvalidation.org/disease/EFO_0000311; Koscielny et al., 2017). A total number of 52 druggable kinases were observed in global proteomic data based on LGG-specific imputed global proteomics table. The association between the global abundance of each kinase and phospho-abundance of phosphosite was assessed via a linear regression. P values were adjusted for multiple comparison via Benjamini-Hochberg adjustment (Benjamini and Hochberg, 1995) and only associations passing a 5% FDR were reported as significant. Then, to assess the enrichment of sites positively associated to a particular module a one-sided fisher-exact test was performed (Table S5).

Pathway analysis of network modules

Gene level pathway analysis was performed for network Module 1 and 2 (referred to as Cluster 1 and 4, respectively in Table S5). Basically, for each network module, we considered the genes whose phosphosites were contained in the network module and identified pathways in the Kegg, Reactome, Hallmark and GO databases enriched in this list. Specifically, a one-sided fisher exact test was performed. Only pathways containing at least 20 genes with phosphorylation measurement were considered for this analysis. Pathways significantly enriched at 10% FDR were found only for Module 1 (Table S5).

Survival analysis of HGG

For this analysis, diagnosis-specific imputed proteomic data was used. There were 25 HGG samples which included 3 patients who had two tumor samples at different time points. For these 3 patients, we used the sample of the initial CNS tumor or the one with the smaller age at specimen diagnosis (if both tumors from the same patient were labeled as progressive/recurrent). Furthermore, 3 HGG samples of autopsies were removed from the analysis. This filtering resulted in 19 HGG samples which were utilized for the survival analysis. Out of these 19 samples, 7 were H3 mutants (all deceased); while the remaining 12 patients (4 alive and 8 deceased) were H3 wild-type. Note that the overall survival was truncated at 2000 days (roughly 5 years) and we treated patients with survival time longer than 2000 days as censored. In particular, only one sample had overall survival greater than 2000 days. This totaled 5 censored samples and 7 "deceased" samples in the H3 wild-type group. Considering these samples, survival data was modeled via Cox regression as follows:

$$\mathrm{Coxph}(\mathrm{OS},\ \mathrm{status}) \sim \mathrm{H3}^{\mathrm{mut}} + \ \mathrm{age} + \mathrm{gender} + \mathrm{post\_treatment} + \mathrm{tumor\_purity} + \mathrm{prot} * \mathrm{H3}^{\mathrm{mut}} + \ \mathrm{prot} * \mathrm{H3}^{\mathrm{WT}}$$

with status denoting the overall survival status. $H3^{mut}$ was coded as one for H3 mutant and zero for wild-type. Gender was coded as 1 for male and 0 for female. Treatment status was coded as 1 for "post-treatment" samples and 0 otherwise. The last two terms in the model denote the interaction between protein abundance and H3 mutant and H3 wild-type, respectively. In particular, $H3^{WT}$ was coded as 1 for H3 wild-type and 0 otherwise. Given that for some HGG tumors, transcriptomic data was not available, for this analysis tumor purity was derived based on global proteomic data via TSNet (Petralia et al., 2018; Table S2).

For 18 of the 19 samples, gene expression data was measured, and we performed a parallel Cox regression analysis based on RNA expression of IDH genes. We derived a 90% confidence interval of hazard ratio estimates for IDH1/2/3 genes based on both global proteomic and gene expression data (Figures S6A and S6B).

To obtain the effect of IDH1 and IDH2 on survival in the H3 wild-type group, we modeled the survival data as function of both IDH1 and IDH2 expression conditional on other covariates as follows:

$$\mathrm{Coxph}(\mathrm{OS},\ \mathrm{status}) \sim \mathrm{H3}^{\mathrm{mut}} + \ \mathrm{age} + \mathrm{gender} + \mathrm{post\_treatment} + \mathrm{tumor\_purity} + \mathrm{IDH1}_{\mathrm{pro}} * \mathrm{H3}^{\mathrm{mut}}$$

$$+ \mathrm{IDH2}_{\mathrm{pro}} * \mathrm{H3}^{\mathrm{mut}} + \mathrm{IDH1}_{\mathrm{pro}} * \mathrm{H3}^{\mathrm{WT}} + \mathrm{IDH2}_{\mathrm{pro}} * \mathrm{H3}^{\mathrm{WT}}$$

For assessing the association between the joint effect of IDH1 and IDH2 proteins on overall survival in the H3 wild-type group, we performed an anova test to compare the above Cox model with the following one:

$$\text{Model 0}: \ \mathrm{Coxph}(\mathrm{OS},\ \mathrm{status}) \sim \mathrm{H3}^{\mathrm{mut}} + \ \mathrm{age} + \mathrm{gender} + \mathrm{post\_treatment} + \mathrm{tumor\_purity} + \mathrm{IDH1}_{\mathrm{pro}}\mathrm{H3}^{\mathrm{mut}} + \mathrm{IDH2}_{\mathrm{pro}}\mathrm{H3}^{\mathrm{mut}}$$

Let the absolute value of the estimated coefficients of $IDH1_{pro}$*$H3^{mut}$ and $IDH1_{pro}$*$H3^{WT}$ be m_1 and w_1 while those for $IDH2_{pro}$*$H3^{mut}$ and $IDH2_{pro}$*$H3^{WT}$ be m_2 and w_2. We calculate the weighted score of IDH1 and IDH2 for the mutant samples as

$$\mathrm{IDH1/2}^{\mathrm{mut}}_{\mathrm{pro}} * \mathrm{H3}^{\mathrm{mut}} = \frac{m_1}{m_1 + m_2} * \mathrm{IDH1}_{\mathrm{pro}} * \mathrm{H3}^{\mathrm{mut}} + \frac{m_2}{m_1 + m_2} * \mathrm{IDH2}_{\mathrm{pro}} * \mathrm{H3}^{\mathrm{mut}}$$

and similarly for the H3 wild-type samples as

$$\mathrm{IDH1/2}^{\mathrm{WT}}_{\mathrm{pro}} * \mathrm{H3}^{\mathrm{WT}} = \frac{w_1}{w_1 + w_2} * \mathrm{IDH1}_{\mathrm{pro}} * \mathrm{H3}^{\mathrm{WT}} + \frac{w_2}{w_1 + w_2} * \mathrm{IDH2}_{\mathrm{pro}} * \mathrm{H3}^{\mathrm{WT}}$$

In order to display the association between survival and weighted IDH1/2 scores in the H3 wild-type group, Kaplan-Meier curves were derived based on $IDH1/2^{WT}_{pro}$ (Hänzelmann et al., 2013), with median value chosen as the cut-off to stratify samples in higher and lower abundance groups (Figure 6E).

We also performed Cox regression to evaluate the association between weighted score of IDH1 and IDH2 with survival conditional on other covariates as follows:

$$\mathrm{Coxph}(\mathrm{OS},\ \mathrm{status}) \sim \mathrm{H3}^{\mathrm{mut}} + \ \mathrm{age} + \mathrm{gender} + \mathrm{post\_treatment} + \mathrm{tumor\_purity} + \mathrm{IDH1/2}^{\mathrm{mut}}_{\mathrm{pro}} * \mathrm{H3}^{\mathrm{mut}} + \ \mathrm{IDH1/2}^{\mathrm{WT}}_{\mathrm{pro}} * \mathrm{H3}^{\mathrm{WT}}$$

We derived 95% confidence interval of the Hazard ratio estimate and other covariates based on the above model (Figure 6C).

We also assessed the association between wild-type IDH1/2 proteins with survival using a second proteomic dataset containing 41 pediatric and young adult HGG patients without IDH1/2 mutants. Among the 41 samples, 12 samples were H3 mutant: 2 alive and 10 deceased. And the remaining 29 samples (19 deceased and 10 alive) were H3 wild-type (Table S6). For survival analysis, we truncated the OS at 2000 days and treat samples with OS greater than 2000 days as censored samples. This left 17 samples with the deceased status in the H3 wild-type group. Similar to the discovery dataset, Cox regression models were fitted on this second dataset, with tumor location further included as a covariate. We used indicators for "cortical" and "midline" tumor location, while cerebellum was taken as the reference. Given that only few markers were available for this dataset, we were unable to derive tumor purity

and include it as a covariate in the model. The KM curve based on weighted IDH1/2 score is displayed in Figure 6F and the 95% confidence interval of the hazard ratio of weighted score in H3 wild-type and other covariates is displayed in Figure 6D.

For pathway enrichment analysis in the pediatric cohort, we used the canonical and Hallmark database from Broad Institute's molecular signature database (Liberzon et al., 2015). We performed a Wilcoxon test to compare the distribution of signed p values (from Cox regression analysis) of the genes within the pathways to the remaining genes in the dataset. We further consolidated the pathways into modules using Sumer (Savage et al., 2019; Figure S6C). Note that we only report the pathway enrichment results from HGG wild-type as this group has reasonably higher sample size as opposed to the mutant group (Table S6).

Drug Connectivity Analysis for HGG

For the transcriptional connectivity analysis, an HGG-specific signature was first generated by comparing the mRNA levels between HGG and LGG samples using the Wilcoxon rank sum test. Genes with an FDR < 0.05 were considered differentially expressed and were subsequently filtered for probes measured in the L1000 assay (Subramanian et al., 2017). The resulting gene list was then used as input for iLINCS, a drug connectivity tool (Pilarczyk et al., 2019) and the "Perturbagen connectivity analysis" functionality was used to identify compounds with negative connectivity to the HGG-specific signature.

For the phosphoproteomic connectivity analysis, protein and phosphopeptide signatures were calculated by comparing HGG and LGG samples via the Wilcoxon rank sum test. Significant phosphopeptide and protein probes (FDR < 0.05) were then mapped to the P100 peptide probes (Litichevskiy et al., 2018) and were used for subsequent analysis. Level 4 P100 data were downloaded from the LINCS Data Portal (Stathias et al., 2020) and the median of each technical replicate was used to calculate the spearman correlation between each P100 experiment and the HGG-specific phosphoproteomic signature. The resulting connectivity scores were then aggregated to the compound level, by calculating the mean among all 7 P100 cell lines. To identify drug MOAs (Mechanisms of Action) that were enriched in the transcriptional and phosphoproteomic connectivity analysis we utilized the fgsea R package (Korotkevich et al., 2019) by querying against MOA drug sets rather than gene sets. Results for multiple data types are included in Table S6.

Comparison of initial and progressed tumors

Our dataset contained proteogenomic profiles of 18 pairs of samples from the same patients. Out of these 18 pairs, 13 of the primary tumors were from the initial disease occurrence, while in the remaining five cases the primary available sample was from a disease that is already classified as progression. Among the secondary samples, 11 are classified as disease progression and seven as recurrence. We analyzed the 18 pairs as cases of less advanced versus more advanced disease, and usually refer to them as initial versus recurrent samples. The mutations included in overlap analysis were the non-synonymous mutations in protein coding genes. Potential driver mutations were either genes with known roles in cancer (Bailey et al., 2018; Li et al., 2015, 2018b) that were found to be mutated, or the ones whose allele frequency increased sufficiently between initial and recurrent samples to indicate the signs of evolutionary selection (Merlo et al., 2006). Chromosome arm copy number activity was defined by comparing arm-level CNVs with their standard deviation across all samples. In particular, arm level amplification was declared if arm-level CNVs were over 1-fold SD above zero, while deletion if arm-level CNVs were more than 1-fold SD lower than zero.

Pathway score differences were computed by subtracting the ssGSEA score (Hänzelmann et al., 2013) of a given pathway in the initial sample from that of the recurrent sample.

Germline variants in TP53

To screen for germline *TP53* variants that are likely to be pathogenic to or causing Li-Fraunemi syndrome, we checked WGS data fromblood/normal samples in CBTN (n = 893). After filtering, we kept germline variants that were either reported before within Li-fraumeni syndrome patients in the literature according to professional version 2019Q2 of the Human Gene Mutation Database (HGMD)® (Stenson et al., 2017), or predicted to be deleterious in *TP53*, which are defined to be i) in the exonic/splicing region, ii) not synonymous SNVs, and iii) with minor allele frequency < 0.001 in both of the gnomAD exome and genome databases (version 2.1.1) (Karczewski et al., 2019). Finally we obtained 19 *TP53* variants in 19 CBTN patients' germline WGS data.

ADDITIONAL RESOURCES

Heatmap Web Server

We have developed a web application (http://pbt.cptac-data-view.org/) which allows researchers to render interactive heatmaps of genes of interest across the cohort, allowing deeper exploration of trends among multiomic and clinical data. The underlying data consists of quantitative information on mutation status, protein abundance, RNA-Seq gene expression, copy number variation, and phosphosite expression for 218 samples when available. The portal has several views available, depending on the data types that the user would like to explore. These views include "all," "mutation," "rna," "proteo," "cnv," and "phospho."

The "all" view provides a multiomic view across multiple data types. Data tracks for each gene are labeled with the gene symbol followed by: "mut"–("Yes" for any type of mutation, "No" for wild-type), "rna"–standardized gene expression levels, or "proteo"–standardized gene-level protein abundance. The "mutation," "rna," "proteo," "cnv," and "phospho" views visualize the individual data tracks. The "phospho" view appends the gene name with a truncated identifier with the amino acid location of the phosphosite, and the user can click the track to see the entire phosphosite identifier.

Please cite this article in press as: Petralia et al., Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer, Cell (2020), https://doi.org/10.1016/j.cell.2020.10.044

All views display the genomic and clinical annotation data as the top tracks. These tracks include survival status, grade, diagnosis, tumor location, and clustering analysis results, including an immune cluster assignment for each sample. The genomic annotation tracks include the mutation status for key genes, including *BRAF* status for LGG, *RELA* for EP, *CTNNB1* for CP, and H3F3A for HGG .

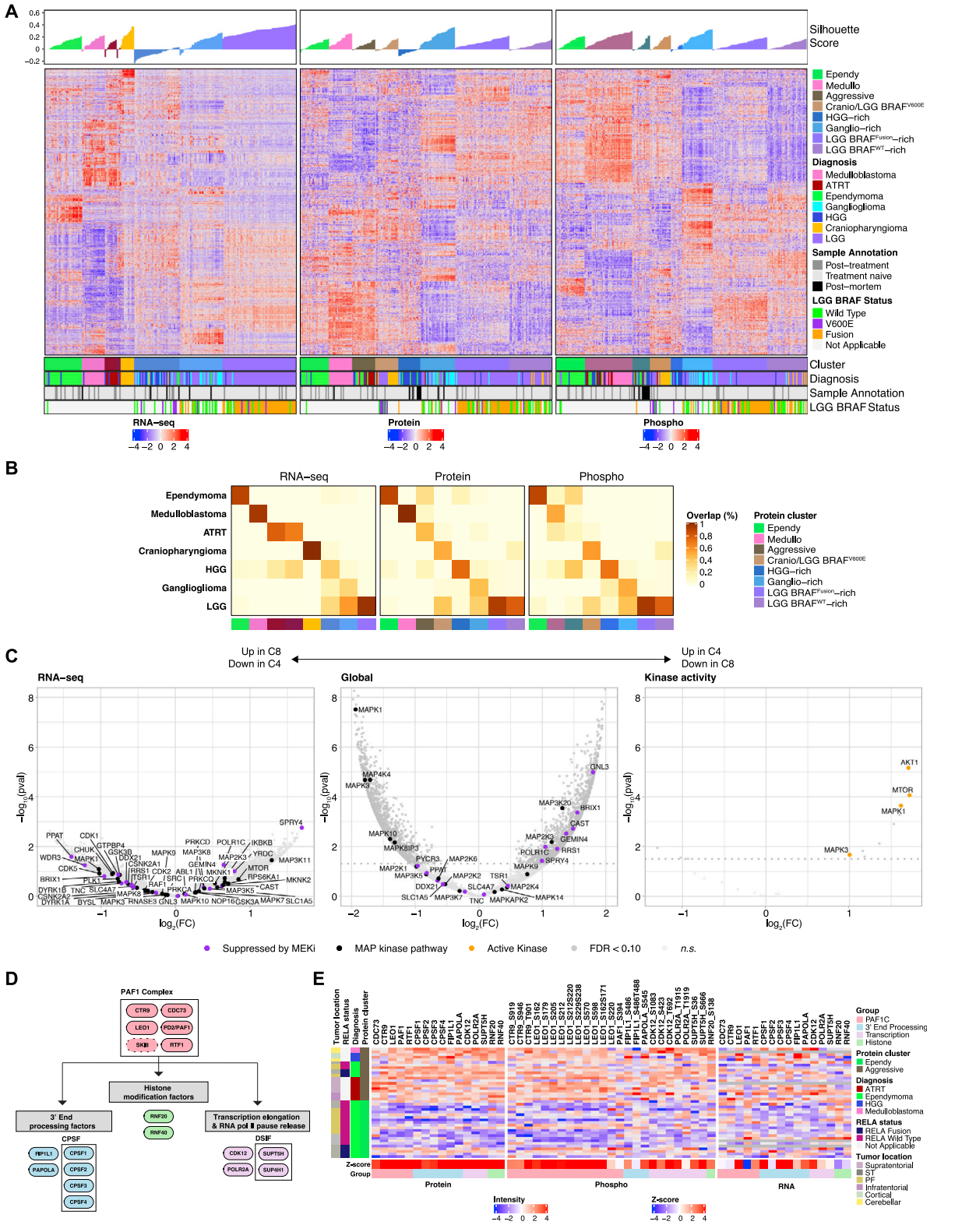
The views can show data for the samples across all histological diagnoses or one individual diagnosis (i.e., EP, MB, ATRT, CP, HGG, ganglioglioma, and LGG).

The application can be accessed with any modern web browser through the following address: http://pbt.cptac-data-view.org. Users begin with a text field, where they can enter gene symbols for up to 30 genes. The genes will be used to generate an Excel file (.xls) and heatmap visualizations across all of the views.

Users can click any point on the interactive heatmap to view the underlying data, including sample identifier, data type, and value. They can then sort the heatmap by a given data track, in ascending or descending order. The sorting feature allows researchers to dynamically explore relationships and patterns among various molecular and clinical data types.

CellPress
OPEN ACCESS

Supplemental Figures



(legend on next page)

Figure S1. Multi-omics-Based Clustering of Pediatric Brain Tumors, Related to Figure 1

A. Clusters based on different omics data (from left to right: RNaseq based, proteomic based and phosphoproteomic based) and corresponding Silhouette scores. For each heatmap, proteomic based clusters (Cluster), different histologies (Diagnosis), sample annotation information and LGG BRAF status are annotated at the bottom of the heatmap.

B. Comparison between proteomic clusters (columns) and histologies (rows). For each histology (rows), the percentage of samples allocated to each cluster (column) is shown.

C. Volcano plot showing genes differentially expressed between C4 and C8 proteomic clusters in CP based on different data types (i.e., RNA-seq, global proteomics, and kinase activity).

D. Diagram illustrating proteins members of the PAF1 complex (SKI8 was not observed in the dataset) as well as downstream players interacting with PAF1C.

E. RNA and global/phospho protein abundance of markers belonging to and interacting with the PAF1 complex based on proteomic and RNA data for EP tumors allocated to the *Aggressive* and the *Ependy* clusters. Proteomic clusters, diagnosis, RELA status and tumor location are annotated on the left of the heatmap. For each gene, the z-score for the comparison between *Aggressive* and *Ependy* clusters is reported.

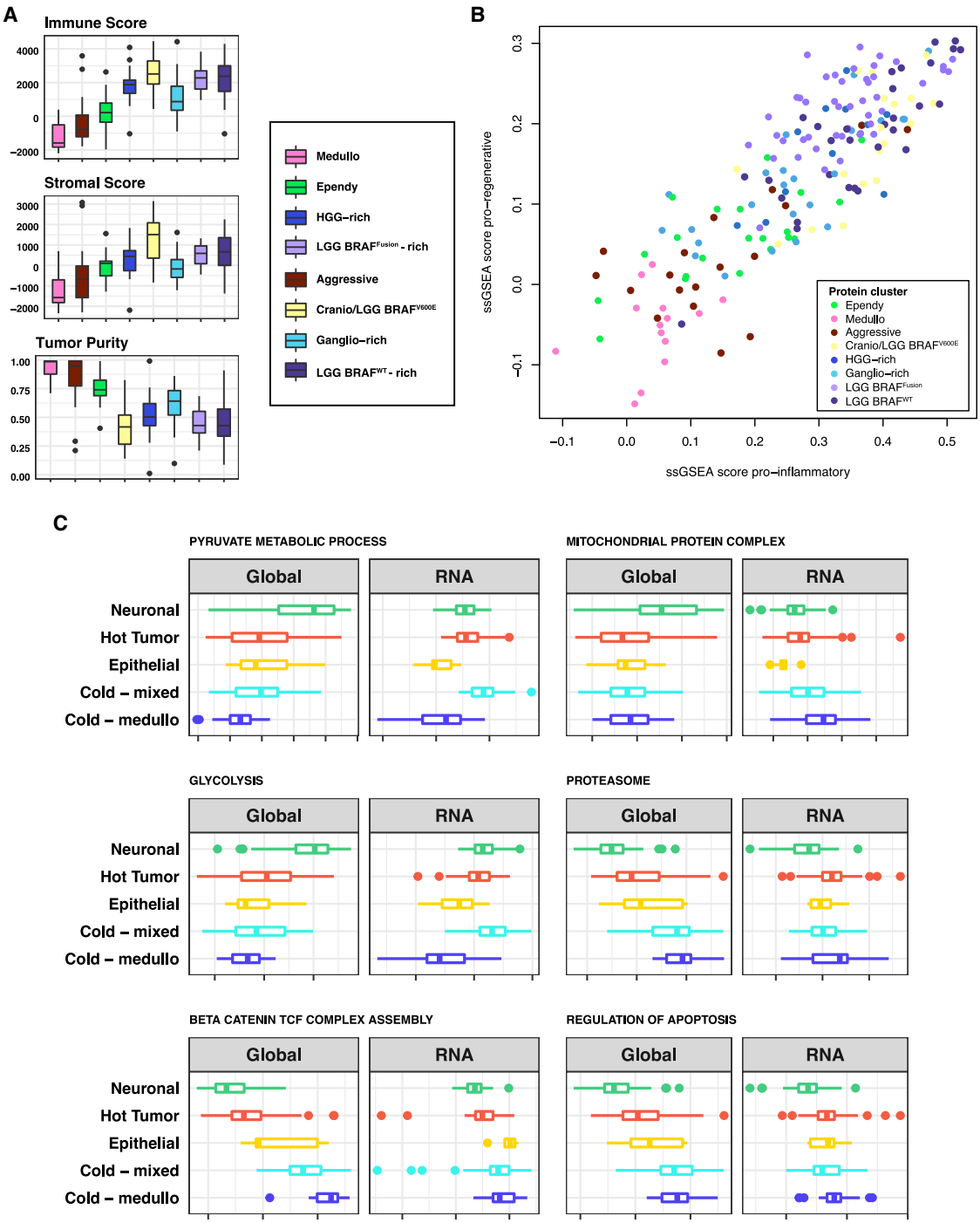


Figure S2. Immune Infiltrations in Pediatric Brain Tumors, Related to Figure 2

A. Distribution of immune and stromal scores from ESTIMATE (Yoshihara et al., 2013), as well as tumor purity estimates from TSNet (Petralia et al., 2018) across different proteomic clusters.

B. Scatterplot of ssGSEA score of pro-regenerative microglia gene signature (y axis) versus that of pro-inflammatory microglia signature (x axis). Colors of the dots represent proteomic clusters.

C. Distribution of pathway scores for Pyruvate Metabolic Process, Mitochondrial Protein Complex, Glycolysis, Proteasome, Beta Catenin TCF Complex Assembly and Regulation of Apoptosis across different immune groups based on RNA and global proteomic data (Global).

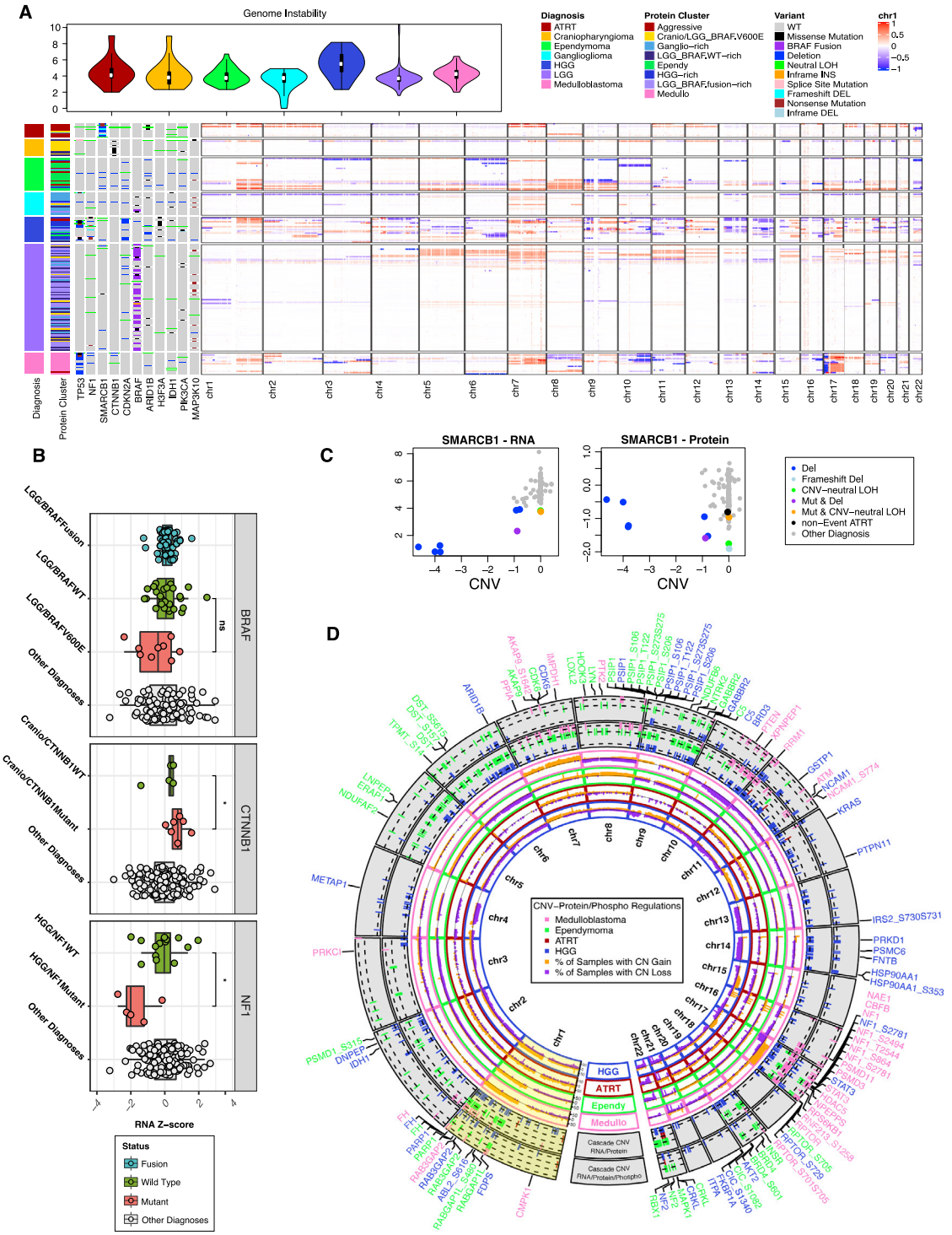


Figure S3. Genomic Alterations and Their Association with mRNA, Protein, and Phosphoprotein Abundances, Related to Figure 3

A. Top: Violin plots showing the distribution of genome instability (log2 scale) for different diagnoses; Bottom-Left: Oncoprint showing mutations in *BRAF, CTNNB1, TP53, SMARCB1, ARID1B, H3F3A, NF1, IDH1, PIK3CA, MAP3K10* and *CDKN2A* across all samples. Bottom-Right: Heatmap showing CNV landscape for all samples.

B. Distribution of gene expression of BRAF, CTNNB1, and NF1 across tumor samples stratified by different mutation status and diagnoses. Symbol * correspond to p values less than 0.1. ns, not significant (p value > 0.1).

(legend continued on next page)

C. Scatter-plot of CNV versus gene expression (left panel) and protein abundance (right panel) of SMARCB1 in ATRT and non-ATRT samples. Colors represent different alteration categories.

D. The four inner circles illustrate copy number amplification and deletion frequencies among HGG, ATRT, EP and MB samples along the genome. Orange bars are for amplifications, while purple bars are for deletion. The outer two circles show the genome locations of diagnosis specific CNV-RNA/protein cascade genes and CNV-RNA/protein/phospho cascade genes respectively. Druggable targets and oncogenes among these cascade genes are further annotated with gene symbols, whose colors represent the diagnoses for which the cascade events were detected.

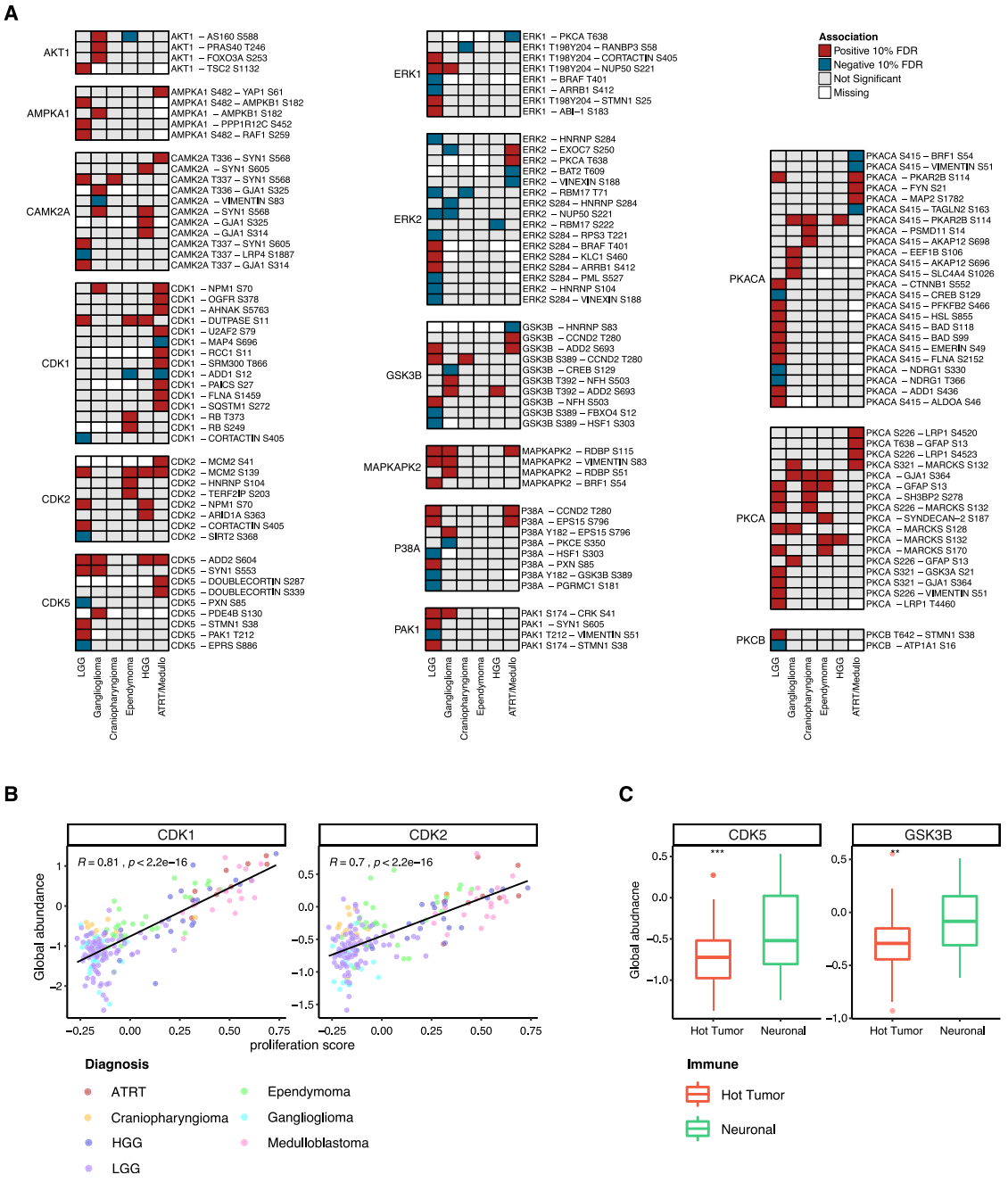


Figure S4. Phosphoproteomics Analysis of Kinase Activity, Related to Figure 4

A. Heatmap showing significant associations between the global/phospho abundances of kinases and phosphosite abundances of substrates among different diagnoses for experimentally validated kinase-substrate interactions from PhosphositePlus (Hornbeck et al., 2015). Kinases are labeled on the left side, while targeted substrates on the right side. Only associations significant at FDR 10% are reported. Positive associations are shown in red, negative associations in blue, and non-significant in gray. For each histology diagnosis, associations were only assessed for sites and kinases observed in more than 50% of the tumors samples of this diagnosis. For sites not passing this threshold within a particular diagnosis, a white cell is shown. To derive these associations, either the global-proteomic or the phospho-proteomic abundances of a kinase are utilized. When the phospho-proteomic abundance is utilized, the name of the phosphosite of the kinase is annotated at the right-side of the heatmap.

B. Scatterplot showing the association between the global abundances of CDK1 or CDK2 (y axis) and the proliferation index (x axis). For each scatterplot, dots are colored based on different histology diagnoses.

C. Boxplot of global abundances of CDK5 and GSK3B for low-grade gliomas stratified by *Neuronal* and *Hot* immune clusters. P values from Wilcoxon-test are reported (i.e., ** corresponding to p value < 0.01 and *** to p value < 0.001)

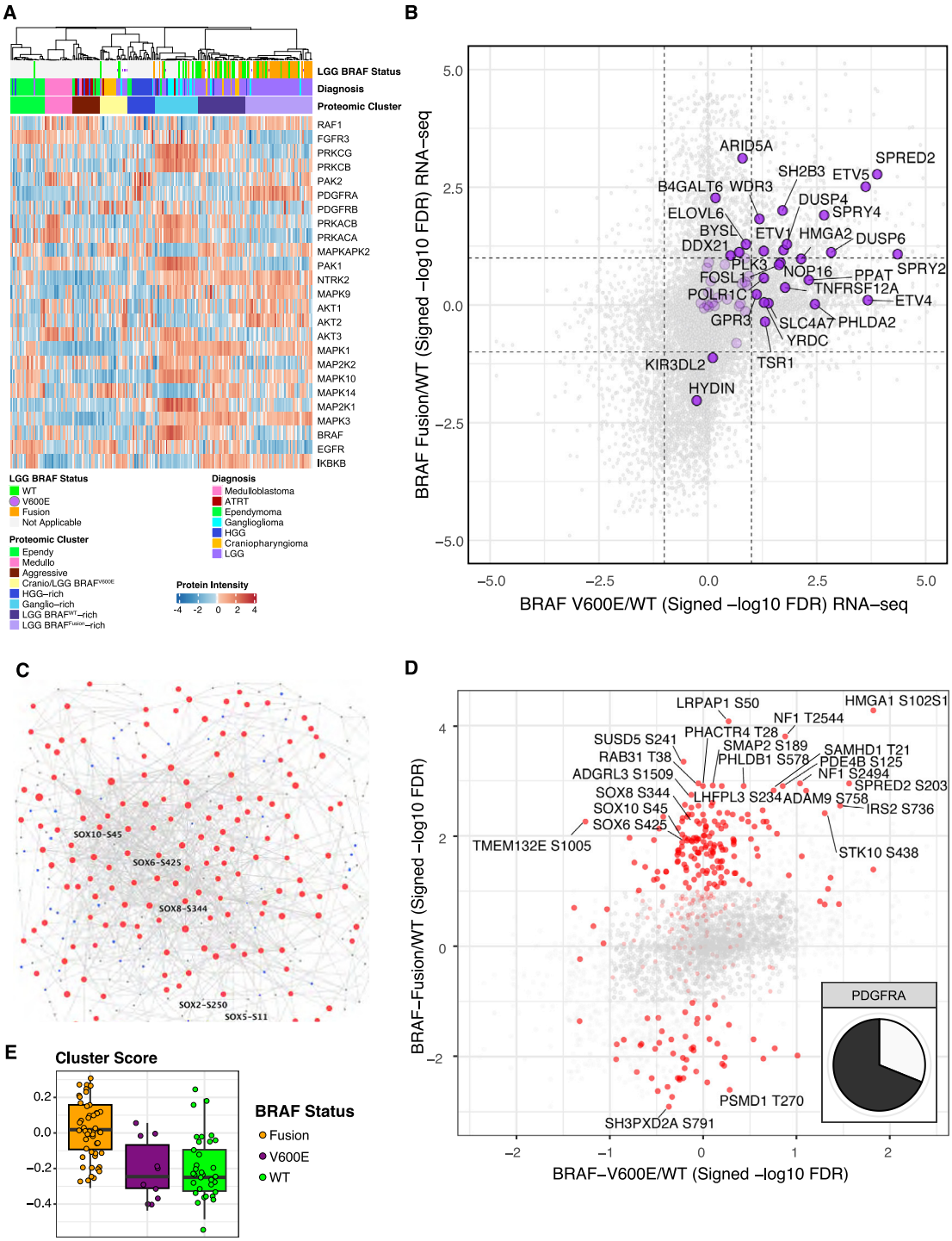


Figure S5. BRAF Status Association and Co-expression Networks Based on Phosphorylation Data of LGG, Related to Figure 5

A. Heatmap of global abundance of key kinases in the MAPK signaling pathway across pediatric brain tumors. Different histologies, proteomic clusters and *BRAF* status (i.e., $BRAF^{V600E}$, $BRAF^{Fusion}$ and $BRAF^{WT}$) are annotated on top of the heatmap.

B. Signed Benjamini Hochberg's adjusted p values (-log10 scale) for the comparison of gene expression levels between of $BRAF^{V600E}$ ($BRAF^{Fusion}$) with $BRAF^{WT}$ tumors are reported on the x axis (y axis). Gene symbols are annotated for genes from the MEK inhibitor signature (Pratilas et al., 2009).

C. The network topology representing the LGG phosphosite co-expression network module enriched with sites upregulated in $BRAF^{Fusion}$ compared to $BRAF^{WT}$ tumors. Nodes correspond to phosphosites while edges correspond to significant associations between phosphosites. Phosphosites positively associated with $BRAF^{Fusion}$ at FDR 10% are displayed in red with node size proportional to the -log10 FDR of the association with $BRAF^{Fusion}$.

(legend continued on next page)

D. Scatterplot of -log10 FDR for the associations between $BRAF^{Fusion}$ (y axis) and $BRAF^{V600E}$ (x axis) with $BRAF^{WT}$. Phosphosites contained in the network module of panel C are highlighted with red. The pie-plot shows the proportion of sites in the network module whose phospho-abundance is associated with the protein abundance of PDGFRA at 5% FDR.

E. Distributions of ssGSEA scores for phosphosites contained in the network module of panel C stratified by different BRAF statuses.

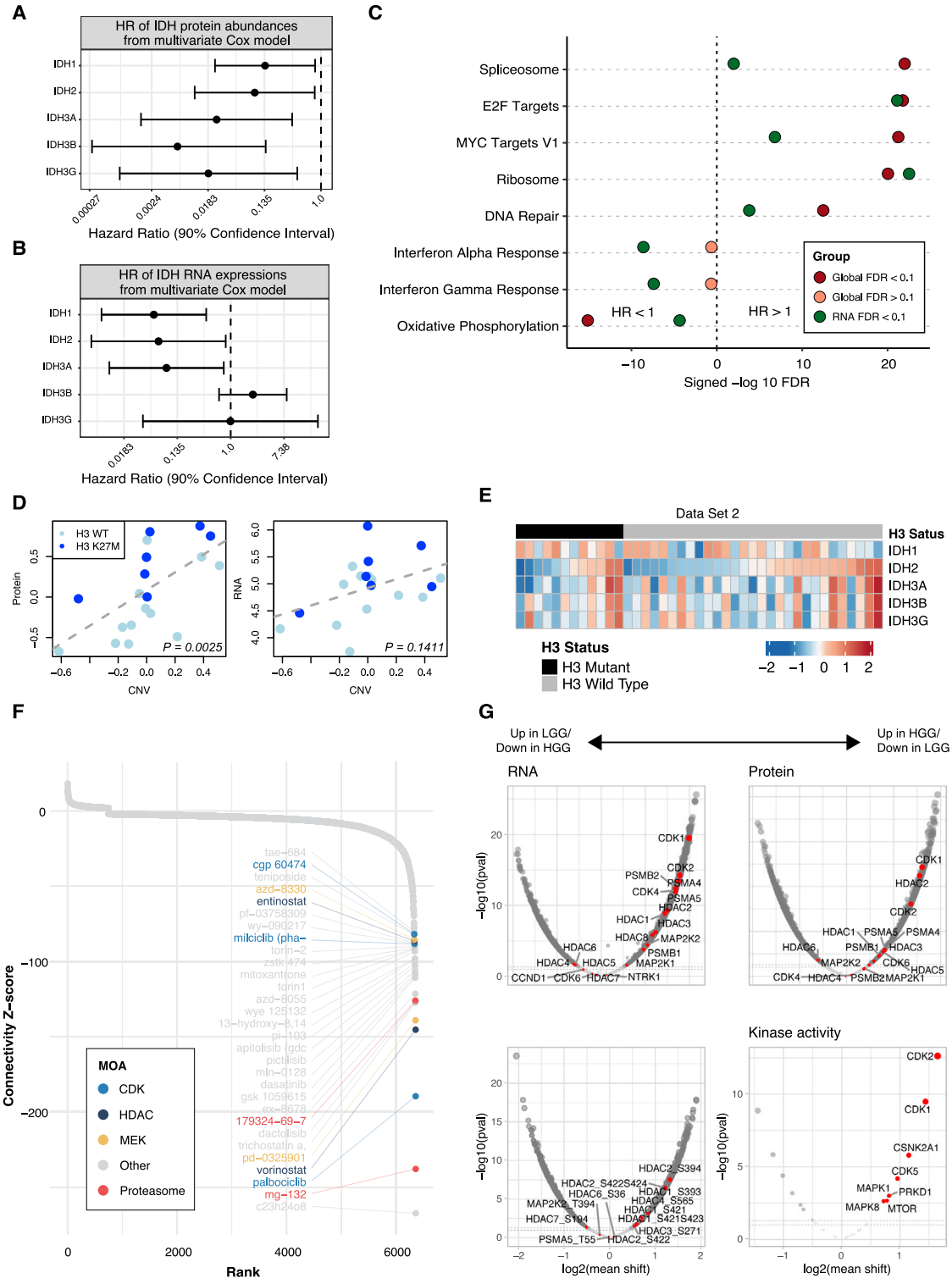


Figure S6. Survival and Drug Target Analysis for HGG, Related to Figure 6

A. Confidence intervals (90%) of hazard ratio coefficients for IDH protein abundances based on multivariate Cox regression models.

B. Confidence intervals (90%) of hazard ratio coefficients for IDH gene expression levels based on multivariate Cox regression models.

C. Pathways associated with survival outcome among *H3*^{WT} HGG patients based on global proteomic (red) and gene expression (green) data. Pathways significant at 10% FDR are marked with darker color.

(legend continued on next page)

D. Scatterplots of the protein abundances or gene expression levels (centered and normalized z-score) versus CNV (log-ratio) of *IDH1* among 19 HGG tumors with CNV data in the discovery cohort.

E. Heatmap of global abundances of IDH1, IDH2, IDH3A, IDH3B and IDH3G proteins in Dataset 2. For each tumor, *H3* mutation status is annotated on the top of the heatmap.

F. Connectivity map score for different drugs based on L1000 Transcriptomics (Subramanian et al., 2017). Different drugs are colored based on the mechanism of action such as CDK1 inhibitor, proteasome inhibitor, HDAC inhibitor and MEK inhibitor.

G. Volcano plots showing genes differentially expressed between high-grade glioma and low-grade glioma tumors based on gene expression, global proteomic, phospho-proteomic data and kinase activity. Genes/proteins annotated are the targets of CDK inhibitor, HDAC inhibitor, proteasome and MEK inhibitors.

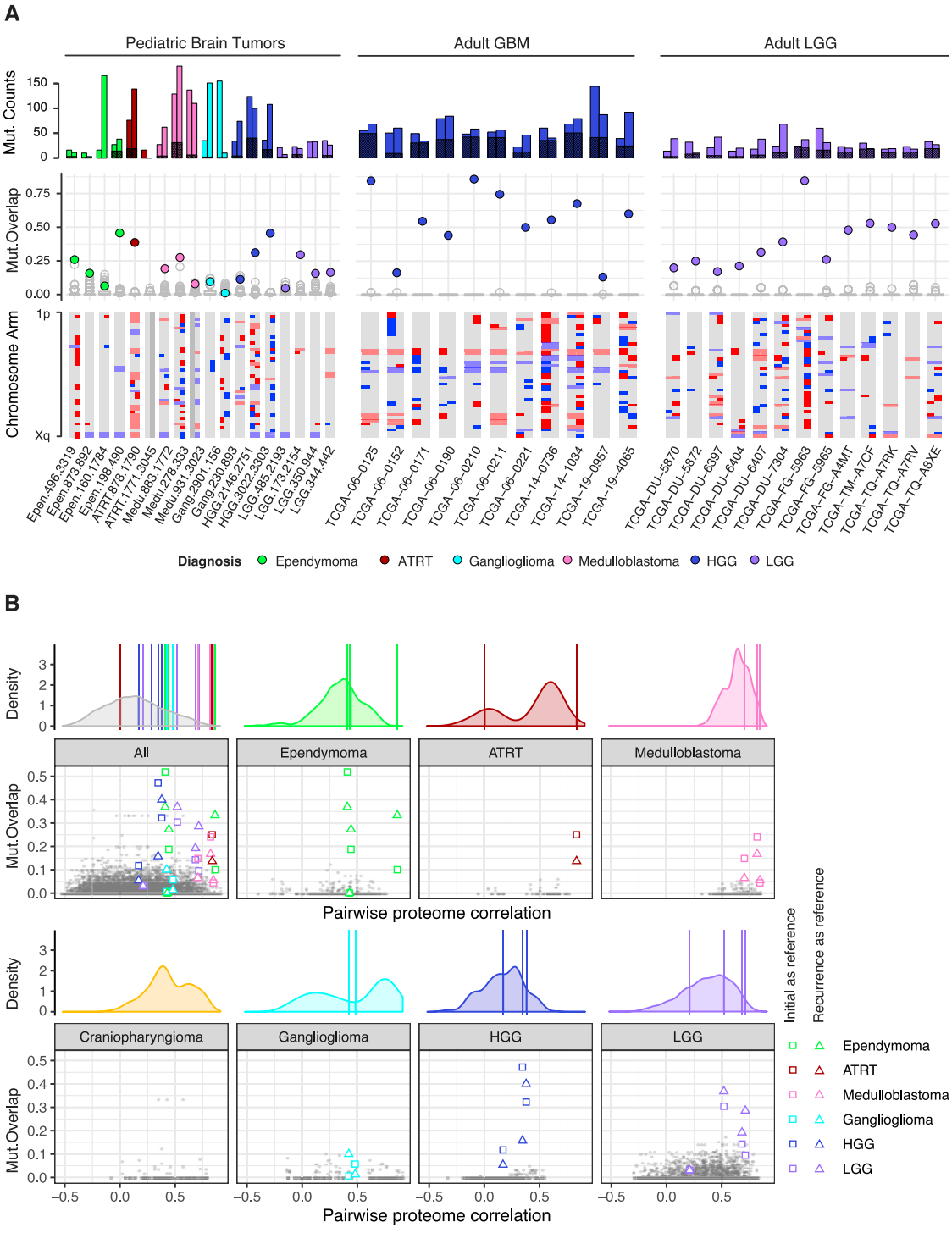


Figure S7. Comparison between Recurrent versus Primary Tumors in Terms of Genomics Alterations and Proteomics Profiling, Related to Figure 7

A. Comparison of mutation counts, shared mutation counts and chromosome arm aberrations between primary and recurrent / progressed tumors in pediatric brain tumors (left), TCGA adult GBM tumors (middle) and TCGA adult LGG tumors (right). Top panels represent mutation counts of paired samples with the proportion of shared mutations highlighted by a shaded area. The middle panels are depicting fractions of shared mutations between each primary tumor and all other tumors from the same dataset, with the recurrence tumor sample of the same patient marked in color denoting the histology. The bottom panels represent significant amplifications and deletions of chromosome arms from 1p to Xq.

(legend continued on next page)

B. Spearman correlations between proteome profiles of tumor sample pairs from the same patients and fractions of mutations that they have in common. The first graph (gray) is for all sample pairs, and the remaining seven (with various colors) are for individual diagnoses. In each graph, the top panel is a distribution of Spearman correlations between all global proteome profile pairs. The values corresponding to 18 primary/recurrent pairs from the same patients are marked with vertical lines. The bottom panels are scatterplots of pairwise sample correlations based on global proteomic abundances versus the fraction of shared mutations. Values corresponding to primary/recurrent pairs are highlighted with colors, where square points represent mutation fractions with reference to the initial tumor sample, and triangles represent mutation fractions computed with reference to the recurrent tumor samples.